Part 1- Session Papers for the EPA 23rd Annual National Conference on Managing Environmental Quality Systems

April 13-16, 2004 Tampa, Florida

Electronic Tools in Quality Systems Management

Automated Audit Software for Streamlining On-Site Laboratory Assessments and Managing On-Going Laboratory Performance (Tim Fitzpatrick, Lab-Data Consultants, Inc.)

Applying Electronic Data Checking Tools to Environmental Data (Wayne Word, EarthSoft, Inc.)

Peer Review as a QA Tool for Photo Interpretation (George Brillis, U.S. EPA)

Project Planning and Assessment

The Transition from DQOs to MQOs for Research Projects (Robert Wright, U.S. EPA)

Greater Consistency of Radiation Emergency Response Analytical Methods Using a Performance Based Approach (Ben Hull, U.S. EPA)

A CD-ROM Based Quality Assurance Project Plan Preparation Tool for Tribes (David Taylor, U.S. EPA)

Electronic Tools for Quality Systems II

Electronic Research Notebooks (Shirley Wasson, U. S. EPA)

Improved Data Quality Systems Using the Florida DEP Automated Date Processing Tool (ADaPT) and EDMS (Juan Manzano, SFWMD)

The Impact of Electronic Record Regulations on the Pharmaceutical and Chemical Industries Maximize Profit or Comply with Regulations? Good Data Management Software lets you do both

Quality Systems Implementation and Assessment

Measures of Quality System Implementation (Louis Blume, U.S. EPA)

Performance-Based Management of Environmental Restoration Program (Javier Santillan, Air Force Center for Environmental Excellence)

Office of Pesticide Program Quality System (Betsy Grim, U.S. EPA)

Environmental Data Quality

What Are Data of Known and Documented Quality? (Marlene Moore, Advanced Systems, Inc.)

Qualifying Quantification of Measurement Variability (Marcus Kantz, U.S. EPA)

Cost-Effective "Collaborative Sampling" in Visual Sample Plan (VSP) Software to Estimate Means and Test Hypotheses (Richard O. Gilbert, Pacific Northwest National Laboratory)

Quality Systems Practices for Research

ETD QA Core Team: An Eloquent Solution to a Complex Problem (Tom Hughes, U.S. EPA)

ETD Quality Control (QC) Checklist for Research Studies (Mette Schladweiler, U.S. EPA)

QA in Research Laboratories: Rules and Reason (Ron Rogers, U.S. EPA)

Automated Audit Software for Streamlining On-site Laboratory Assessments and Managing On-going Laboratory Performance

Tim Fitzpatrick
Laboratory Data Consultants (LDC)
7750 El Camino Real, Suite 2L
Carlsbad, CA

Delia Ivanoff and Larry Teich
South Florida Water Management District (SFWMD)
1480 Skees Road
West Palm Beach, FL

Abstract:
As part of improving the process for performing on-site laboratory audits and maintaining and managing QA/QC documentation in a readily available electronic format, Laboratory Data Consultants, Inc. (LDC) under contract to the South Florida Water Management District (SFWMD) developed a Microsoft ACCESS based software program to streamline the audit preparation, on-site audit process, final reporting, and long-term documentation. The program has two primary components: (1) a master database which contains laboratory information including names of key staff, certification status, performance evaluation (PE) data, past audits and corrective action, SOPs, and a list of methods in which the lab is certified and (2) a "briefcase" database which is downloaded from the master database prior to performing the on-site audit. This "briefcase" database is taken to the on-site audit and contains the laboratory specific audit checklists based on SFWMD, Florida DEP, and NELAC standards.

This presentation will show how the use of this audit software program has made the audit process more consistent, technically sound, cost effective, and real-time for the auditor. The main features of this program include:

- *Guides the auditor step-by-step through a NELAC type audit.*
- *Allows for easy electronic access to all past audits and associated corrective action*
- *Ability to prepare specific questions on the audit checklist based upon past audits.*
- *Access to past SFWMD PE sample results.*
- *Provides e-mail notification of critical dates in the audit process.*
- *Tracks time each auditor has spent on the audit.*
- *Embedded reference documents such as Chapters 3 and 5 of the NELAC standards and the SFWMD QA Manual are available for electronic review during the audit process.*
- *Reference tables that link findings to specific sections of NELAC documents.*
- *Direct input of findings into a laptop computer.*
- *Automated final report generation based upon findings listed during the on-site audit and information retrieved from the master database.*
- *Tracks findings and corrective action responses from the laboratory.*

In summary, the SFWMD Automated Audit software has demonstrated to be an extremely powerful tool in aiding the auditor to be better prepared and perform on-site audits in a

cost effective, technically sound, and consistent manner. Additionally, LDC and the SFWMD are in the process of expanding the software to accommodate field audits.

Introduction

Preparation for an audit has long been a time-consuming process involving the review of multiple documents such as laboratory-specific SOPs, laboratory Quality Assurance Manuals (QMs), internal quality assurance audit results, historical certification audit results, EPA and state-specific analytical methodologies, proficiency and round-robin testing results, training records, raw data, and more. Once the audit process begins, the auditor generates another stack of documents. There are standardized checklists and issue-specific questions/issues to investigate, lists of findings and related corrective actions/recommendations, response from the lab to the auditor's findings, the auditor's response indicating acceptability of the lab's corrective action plan, etc.

The automated audit software developed by Laboratory Data Consultants, Inc. (LDC) under contract to the South Florida Water Management District (SFWMD) addresses the difficulties associated with the preparation for and execution of an on-site audit as well as the sometimes overwhelming task of organization, storage and retrieval of the massive amount of documentation associated with the audit process.

Conducting the Audit

The auditor does all the preparation work in the central database. The auditor is guided step-by-step through the preparation. The first step involves importing relevant documents (QMs, SOPs, previous audits, etc.) into the central database. These documents are either supplied in electronic format by the lab or scanned and converted to pdf files for electronic storage and retrieval.

Once the database is populated with these documents, the auditor begins the process of building an electronic "briefcase" which will be downloaded to a laptop computer and taken to the on-site audit. The auditor tags the documents and references to be imported and can hyperlink these documents to questions in the checklist. The auditor can then open any of the references during the course of the audit by simply clicking on the embedded hyperlink.

The auditor next selects the analytical and prep methods to be audited from a list of the more common test methods in use by SFWMD. Methods not on the list can be easily added by the auditor through a wizard-style interface.

The standard checklist consists of questions based on SFWMD, FDEP and NELAC requirements and regulations. The standard checklist can be easily appended by the auditor in several categories including previous audit results, performance testing results, internal audit results and miscellaneous. Questions can be added either during the preparation in the central database or in the briefcase during the course of the audit.

The program also contains a schedule and timeline tracker which is accessible from both the

central and briefcase modules to keep track of critical dates in the audit process. The scheduler will send a reminder email to the auditor one month in advance of scheduled audits. A timelog is also embedded in the product so that the auditor can keep track of the hours spent on preparation and execution of the audit.

After the auditor has selected the lab, methods, reference documents and any additional questions, the entire audit package is then saved as a "briefcase" database. The auditor next imports this file into the briefcase module of the software which is normally on a laptop computer. The auditor will take this to the on-site audit where he can work either directly from the electronic forms or from a printed hardcopy. Additional pertinent documents such as run logs, training records, raw data, etc. may also be added to the briefcase as pdf files during the course of the audit.

A tabular summary of findings and an audit assessment report are generated based on responses to questions in the audit checklist and the opening meeting worksheet. The summary of findings is then immediately available for distribution to the lab staff at the closing meeting. This tabular summary can be exported in MS Excel format as well.

The auditor msy add freeform conclusions and make minor modifications to the assessment report; however, use of the template both streamlines the preparation and standardizes the format of the assessment report.

The report is sent to the lab in both hardcopy and electronic formats. The laboratory's response is then imported and a letter of acceptance or further action required is generated from a template. This process is repeated as necessary until resolution of all issues is complete. Once completed, the briefcase database is then uploaded back to the central database for archiving, virtually eliminating the need for paper filing and storage.

Conclusion

In summary, the SFWMD Automated Audit software is an extremely powerful tool in aiding the auditor to be better prepared and perform on-site audits in a cost effective, technically sound, and consistent manner. The software can be used to perform both internal and external audits. LDC is currently in the process of developing a module for conducting field and quality systems audits as well.

Attachments

A series of screenshots from a mock audit follows this section showing details of some of the key screens and features of the software.

Figure 1 – Central Database

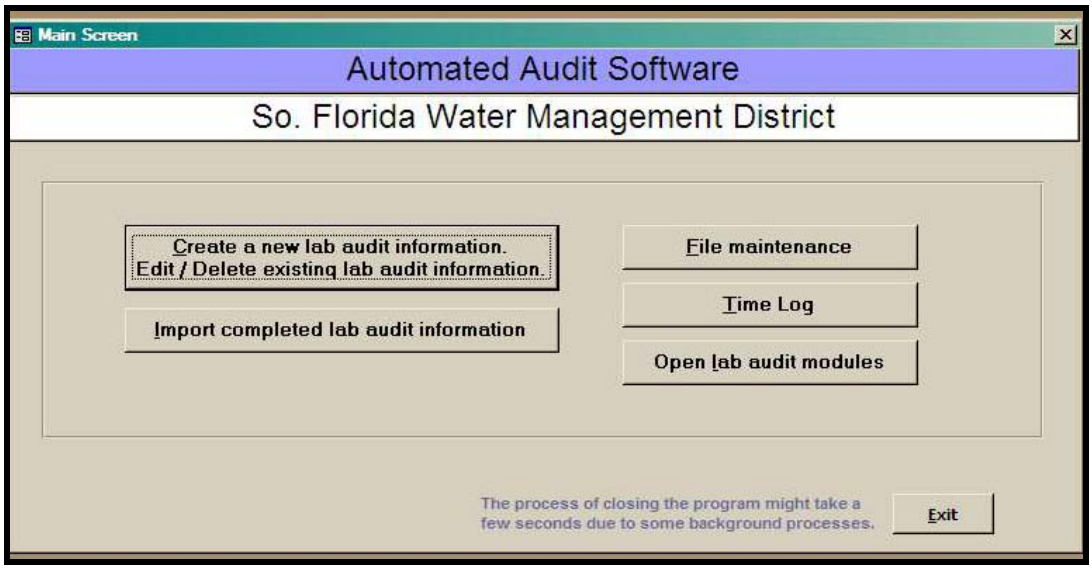


Figure 1 shows the main screen of the Central Database Version. From here, the auditor can choose to begin a new audit, edit or delete an existing audit, upload a completed audit briefcase to the central database or perform various file maintenance functions such as importing documents, adding labs or personnel, modifying the checklist, etc.

Figure 2 - Briefcase

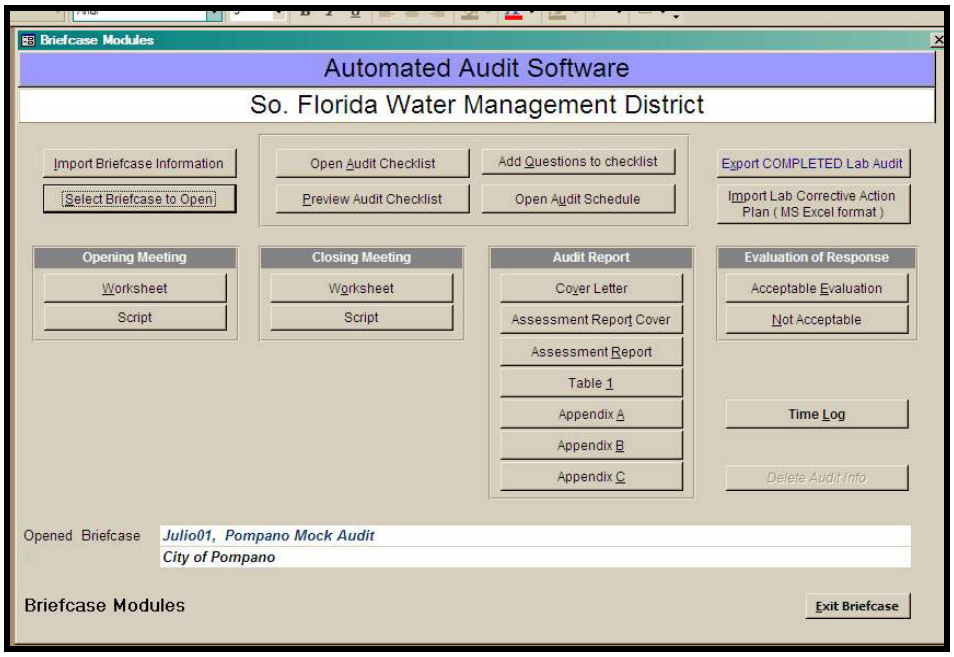


Figure 2 shows the main screen of the Briefcase version. This is where the on-site audit is conducted and reports are prepared.

Figure 3 – Audit Checklist

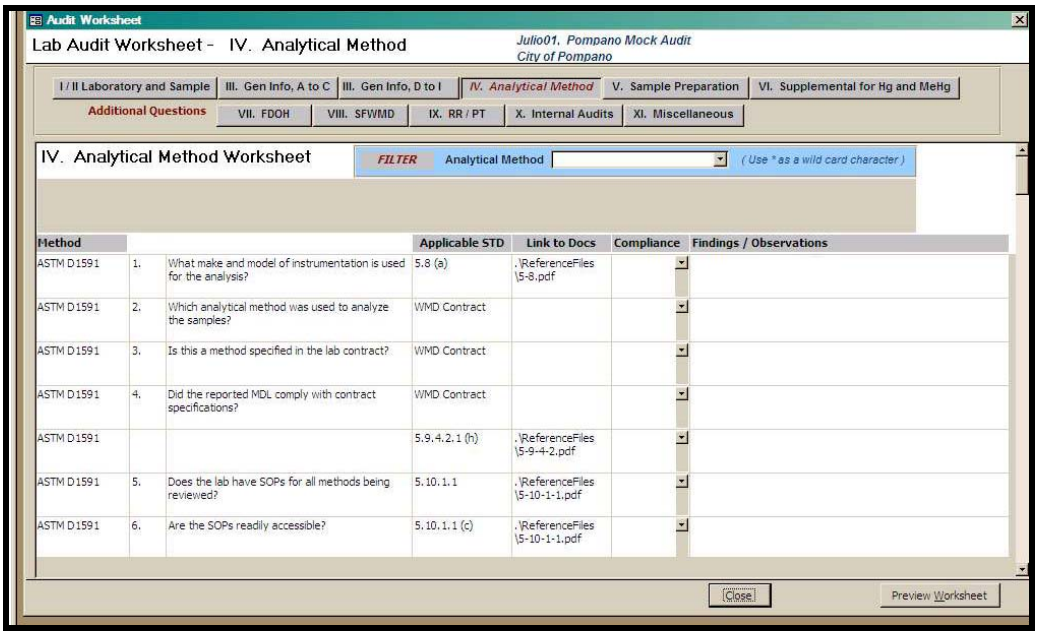


Figure 3 shows a portion of the audit checklist. Separate sections of the list are opened by clicking on the appropriate button. This screenshot shows the analytical method worksheet. This checklist is filled out for each method being audited.

Figure 4 – Opening Meeting Worksheet

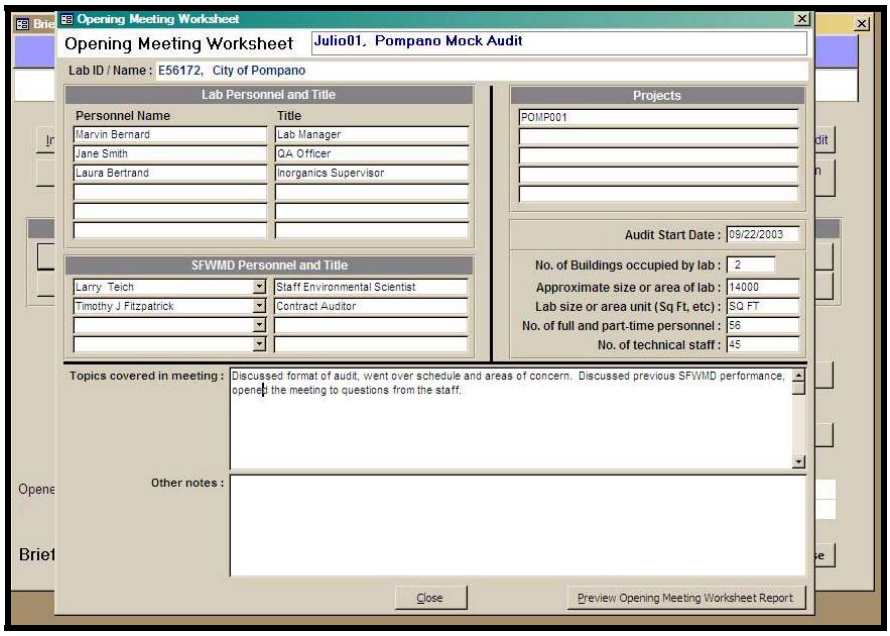


Figure 4 shows the opening meeting worksheet. Information entered here is directly imported into the final assessment report template.

Figure 5 – Assessment Report

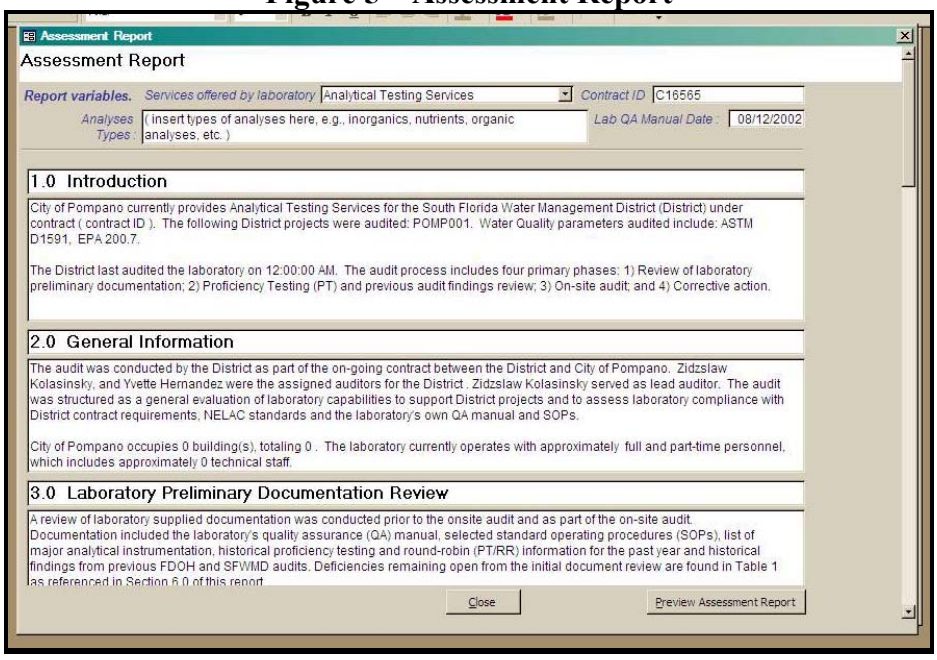


Figure 5 shows the Assessment Report template. The report is prepared automatically based on entries made in the checklists. The auditor may add freeform conclusions and make minor modifications to the text of the template.

Figure 6 – Summary Table

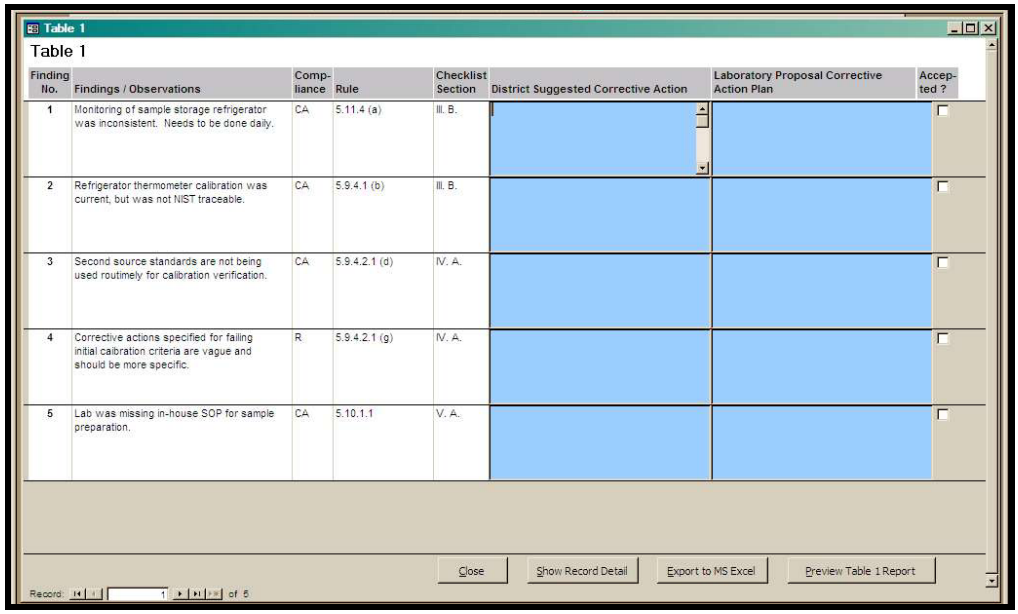


Figure 6 shows the format of the tabular summary of findings and recommendations. Clicking on ***Show Record Detail*** will take the auditor directly to the checklist entry which produced the finding.

Applying Electronic Data Checking Tools to Environmental Data

Wayne Word, EarthSoft, Inc.

Introduction

The rate at which environmental data is being generated today and the importance of managing that data in real time with reduced resources is driving a renewed look by all participants in the value of using electronic tools in the collection of environmental data.

This presentation will introduce the concepts of electronic data download (EDD) file definitions, data checking, conditional processing, and automated data transmittals. A leading edge data PC processing tool will be introduced and opportunity for hands on experience will be provided. Further automation of the data path for data importing and subsequent actions on receipt such as action level criteria checking; push reporting; and automated emails will also be explored. A new generation data verification tool will also be presented that assists data validators in finding the "needle in the haystack" errors that require attention to ensure the data received meets the project's intended data quality objectives.

Additional opportunity to review the products and procedures presented will be afforded following the presentation, or upon request at www.earthsoft.com.

Key Concepts Presented:

- Importance of Data Quality
  - First address **Data Quality**, then **Data Usability**
  - Data Usability without Data Quality is 'Garbage In - Garbage Out'.
  - Bad data can make beautiful graphs! But, Good data leads to Good Decisions.

- Existing Challenges
  - Variations in formats
  - Inconsistent values
  - Missing content
  - Errors in content
  - Verification at end of the process leaves little ability to make corrections – TQM?
  - Much Manual effort required for success

- Solutions
  - Define data formats for each data type
  - Establish consistent valid values
  - Check EDDs for compliance and report errors for correction – at generation
  - Verify EDD at receipt
  - Define data package for automated processing

- EDP – for Data Submitters

- A .Net product, using XML and easily deployed.
- Data producers can ***check*** and ***correct*** EDDs.
- ***Custom Checking*** scripts may be included with formats – special checking can be configured for:
  - Required fields and Reference Values
  - Conditional checking (if this field is X, then that field should be Y)
  - Other special or custom required data reviews QC recovery checking
  - Check ***multiple*** EDD format tables simultaneously

- EQuIS Data Processor (EDP) - Additional Workflow Values
  - Allows end user to use XML to create new EDDs.
  - User Login/identification, and Audit trail capable.
  - Generate error report including row number, error, and value.
  - Screen data against required limits, with different regulatory limits per location.
  - Upload EDD from user, via email, ftp, or portal.
  - Web Enabled to automate EDD processing for high volume operations.
  - Optionally upload data into defined database tables.

- Data Entry - DATA ENTRY FORMS for manual data entry are useful for historical data conversions and for field data entry.
  - This XML tool for Laptops or Tablet PCs enables a controlled dialog to create the import EDD files in a familiar layout for field staff.

- EDP Online
  - Labs and Data Providers can submit data online or via email.
  - Pass or Fail results (with Error Logs, if any) are emailed back to data provider.
  - Automation By Email or Online
    - Verify Submitter
    - Auto process Package
    - Maintain Audit Trail
    - Email EDD messages
    - Load accepted data
    - Initiate Push processing

- DQM – Data Qualification Module
  - Automate the process of Data Verification Checks.
  - Reduce staff time checking good data.
  - Find the Needle-in-a-haystack errors.
  - Focus experts on resolving problems.
  - Enable higher percentage checks – up to 100% for key data items.

- DQM Procedures
  - Review Data by SDG
  - Set Up Criteria by Method/Matrix
  - Enter Method and Matrix Specific Holding Times, Qualifiers, etc…
  - Select Checks and Qualifier Rules

- Run Checks, Review Error Log
- Commit Results to the Database
- Generate Report Tables
- Electronic Checks

- DQM Benefits for Data Verification
  - Produce better quality and more reliable data.
  - Implement discovery and correction procedures at data submittal – TQM.
  - Enable checking of a higher percentage of data.
  - Enable automation of routine processes.
  - Better, Faster, Cheaper overall process.

PEER REVIEW AS A QA TOOL FOR PHOTO INTERPRETATION

George M. Brilis,
John G. Lyon
U.S. EPA Office of Research and Development
National Exposure Research Laboratory
Environmental Sciences Division
P.O. Box 93478
Las Vegas, NV 89193-3478

Jeffrey C. Worthington
U.S. EPA Office of Environmental Information
1200 Pennsylvania Avenue, MW
Washington, DC 20460

The information in this document has been funded wholly or by the United States Environmental Protection Agency. It has been subjected to the Agency's peer and administrative review and has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use. Opinions contained in this article are solely those of the authors and do not reflect the opinions, or the future direction, of the US EPA or the Federal government.

Abstract:

Remotely Sensed (RS) images are used in many ways in the EPA. Eventually, the photo may be "interpreted." When images are interpreted, attempts are made by humans to determine what is on the ground (or in the air) by examining the photo or image, and the implementation of Quality Assurance (QA) takes a unique turn.

Typically, the quality of an analysis is most often assessed by comparison of a product or result to an external, tangible standard. Deviations from that standard may be a reflection of quality.

On the other hand, photo interpretation is, for the most part, a subjective process. Historically, the interpreters's experience and education must be relied upon for quality assessment.

This paper proposes an additional means for quality assessment, namely peer review. The EPA published the EPA Peer Review Handbook, 2nd Edition, EPA 100-B-00-001, in December 2000. The reader is encouraged to review this document for detailed information on peer review.

This proposal combines the EPA "graded approach," QA Categories, and Peer Review in a strategy designed to provide known quality, and ensure that the photo interpretation results are defensible. Unlike QA in other disciplines, photo interpretation is unique in that peer review may be used as a QA procedure, not strictly as a

means to "...determine acceptance by the general scientific community.."[1] (US Supreme Court, Daubert v Dow, 1993).

The Process of Photo Interpretation

Photo Interpretation is not entirely a subjective process. There are basic rules and procedures used in photo interpretation. Most individuals interpret images everyday, such as photographs in newspapers. Photo interpretation experts have developed skills over the years, using some important fundamentals. Portions of the following text are taken from *Remote Sensing and Image Interpretation*, Lillesand & Keifer, Fourth Edition, Wiley, 2004. It is important for the reader of this paper to be familiar with some basic image interpretation fundamentals.

Interpretation of aerial and space images differ in that they are:

1. Portrayal of features from an overhead, often unfamiliar perspective.
2. Frequently use wavelengths outside the visible portion of the spectrum, and
3. Depict the earth's surface at unfamiliar scales and resolution.

These differences are what make photo interpretation unique and are why education and experience are required for accurate photo interpretation. Most interpreters will consider the following basic characteristics when interpreting:

- Shape - the general form, configuration, or outline of individual objects help in their identification
- Size - relative to other objects and also relative to the scale of the image.
- Pattern - relating to the spatial arrangements of objects, for example repetition of a particular form may help interpretation.
- Tone (or hue)- refers to the relative brightness of an object on an image - note that the brightness may change depending on the wavelength being used
- Texture - is the frequency of tonal change on an image - an object may appear "smooth" or "coarse."
- Shadows - aid interpretation by helping discern shape and give valuable information about the objects within the shadow.
- Site - the topographic or geographical location is particularly important in the identification of vegetation types
- Association - refers to the occurrence of certain features in relation to others.
- Resolution - depends on many factors - some objects may be too small or have too little contrast with their surroundings to be clearly seen.
- Other factors, such as image scale, image color balance, and condition of the image (faded or torn) also affect the interpretation.[2]

---

[1] *Quality Science in the Courtroom: U.S. EPA Data Quality and Peer Review Policies and Procedures Compared to the Daubert Factors*, Brilis, Worthington, Wait, Environmental Forensics, Vol.1, No.4, pgs 197-203, December 2000

[2] *Remote Sensing and Image Interpretation*, Lillesand & Keifer, Fourth Edition, pgs 193-194, Wiley, 2004.

The interpreter will also invoke certain strategies when approaching an interpretation. These strategies are logistic and most typically utilize the process of convergence of evidence to arrive at a conclusion. The interpreters themselves may be experts in different fields. For example, vegetation, landscaping, military equipment, water, agriculture, etc.

Again, the readers of this paper are encouraged to review image interpretation literature to familiarize themselves with the practical aspects of photo interpretation. By doing so, one has a better sense of where the subjectivity of interpretation resides, and how education and experience contribute while at the same time limit an interpreters' accuracy.

The "Graded Approach"

The EPA Quality Staff (QS) developed and implemented a "graded approach" philosophy. This philosophy is a common sense approach that "establishes QA and QC requirements commensurate with the importance of the work, the available resources, and the unique needs of the organization."[¼]

QA Categories

EPA's Office of Research and Development (ORD) QA Staff has adapted the graded approach and created four levels of QA categories. The following four categories are generally accepted and recognized by EPA's Scientific and QA Communities[4,5]:

Category 1: Category 1 projects require the most rigorous and detailed QA and QC, since the resulting data, information and conclusions, must be both legally and scientifically defensible. These projects are of sufficient scope and substance that their results could be used directly, without additional support, for compliance or other litigation. Such projects are of critical importance to the EPA's mission and goals and must be able to withstand the rigors of legal challenge. Category I projects include enforcement actions and projects of significant national or congressional visibility. Such projects may be monitored by the EPA Administrator. These projects directly and/or immediately supports specific Agency rule-making, enforcement, regulatory, or policy decisions. The data, information and conclusions of Category 1 projects must be autonomous; that is, a

[3] *EPA Quality Manual for Environmental Programs*, 5360 A1, pp. 3-1, US EPA, May 5, 2000.

[4] *Preparing Perfect Project Plans: A Pocket Guide for the Preparation of Quality Assurance Project Plans*, Guy F. Simes, EPA/600/9-89/087, pgs iii - vi, October 1989.

[5] *Integrated Information and Quality Management Plan*, Rev 0, Betz., Brilis, Holm, Hunike, Johnson, L., Kantor, Martinson, Viebrock, pg 66 & 67, Document Control Number: "NERL IIQMP No. 1," US EPA, Spring 2002.

Category 1 project can prove or disprove a hypothesis without reference to complementary projects.

Category 2. These projects are those producing results that complement other inputs. These projects are of sufficient scope and substance that their results could be combined with those from other projects of similar scope to produce narratives for making rules, regulations, policies, or laws. In addition projects that do not fit this pattern, but have high visibility, would also be included in this category.

Category 3. Category 3 projects produce data, information and conclusions, that are used to evaluate and select basic options, or to perform feasibility studies or preliminary assessments of unexplored areas which might lead to further work. In research, these projects provide demonstration or proof of concepts - this includes method validation studies.

Category 4. These projects produce data, information and conclusions, for the purpose of assessing suppositions. In research, these projects are basic, exploratory, conceptual research to study basic phenomena or issues. This includes the characterization of health or ecological mechanisms and/or endpoints in order to improve the understanding of the interaction of environmental compounds, conditions, or processes with human and other life forms. This also includes development of assays or methods for detecting or estimating the influence of a particular environmental agent on a specified health or ecological endpoint.

In addition to providing the QA and scientific professionals a means of quickly determining the QA requirements of a project, the Categories also provide managers and decision makers with a way to identify the importance of a project. Managers and Decision Makers may also use the categories as a way of ordering work to be conducted.

Peer Review "Categories"

The EPA published a handbook for planning and conducting peer reviews[6]. Though the *Handbook* addresses many of the aspects of peer review, it does not address many details for which specialized expertise exists elsewhere.[7] Similarly, this paper proposes a conceptual solution using peer review in lieu of an objective, tangible standard for quality assessment.

The US EPA has not officially categorized *levels* of peer review, however criteria have been delineated to determine whether or not a project is "major" and should be peer reviewed.

---

[6] *EPA Peer Review Handbook*, 2nd Edition, Deerfield, Flaak, Schumann, Wentworth, Hammett, Kuzmack, Snyder, EPA 100-B-00-001, pgs 26-27, December 2000

[7] *EPA Peer Review Handbook*, 2nd Edition, Deerfield, Flaak, Schumann, Wentworth, Hammett, Kuzmack, Snyder, EPA 100-B-00-001, pg xi, December 2000

EPA also uses the peer review process in the planning stages of a project, prior to the initiation of work.

The EPA defines "major" scientific and technical work products as those that are used to support a regulatory program or policy position and that meet one or more of the following criteria are candidates for peer review:

- Establishes a significant precedent, model, or methodology
- Addresses significant controversial issues
- Focuses on significant emerging issues
- Has significant cross-Agency/inter-agency implication
- Involves a significant investment of Agency resources
- Considers an innovative approach for a previously defined problem/process/methodology
- Satisfies a statutory or other legal mandate for peer review.

"Usually, a major scientific and/or technical work product supports a regulatory decision or policy/guidance of major impact. Major impact can mean that it will have applicability to a broad spectrum of regulated entities and other stakeholders, or that it will have narrower applicability, but with significant consequences on a smaller geographic or practical scale. The scientific and/or technical work that underlies many of the Agency's major rulemakings and policy and guidance documents of general applicability would be designated "major" under this scope of impact criterion because of their far-reaching or significant impacts. The novelty or controversy associated with the work product helps determine whether it is major or not. A major work product may be novel or innovative, precedential, controversial, or emerging ("cutting edge"). An application of an existing, adequately peer reviewed methodology or model to a situation that departs significantly from the situation it was originally designed to address is a candidate for peer review. Similarly, a modification of an existing, adequately peer reviewed methodology or model that departs significantly from its original approach is a candidate for peer review."[©] Determination of "significant departure" as used in this paper is the responsibility of the Decision Maker.

In summary, if a scientific or technical work product has a major impact, involves precedential, novel, and/or controversial issues, or if the Agency has a legal and/or statutory obligation, then a peer review is required.

Currently Accepted Techniques of Validating Photo Interpretation

Recognizing the subjectivity of photo interpretation, the EPA currently uses one or both of the following means of validating the interpretation results:

1. Ground Truthing - This technique requires that a qualified individual physically carry the results to the location and verify the interpretation by examining the surroundings. This technique is not always practicable because of; a) the associated costs of travel; b) the possibility that the location is physically

[8] *EPA Peer Review Handbook*, 2nd Edition, Deerfield, Flaak, Schumann, Wentworth, Hammett, Kuzmack, Snyder, EPA 100-B-00-001, pgs 26-27, December 2000

inaccessible; c) the attributes of the location may have changed since the interpretation took place

2. Informal Peer Review - In this type of peer review, the factors described in the EPA Peer Review Handbook are not always considered. Often, the interpretation is reviewed by management prior to release, with the simple perspective of "does it make sense."

These two techniques, though valuable in their own right, may not withstand the scrutiny that evidentiary materials are given in litigation.

Resources

As stated, Ground-Truthing and Peer review are the most common, and possibly only, tools available for final QA assessment of a photo interpretation. Both are resource-intensive processes.

Ground-truthing

In itself may simply be a matter of driving 0.5 miles to examine a site or require a helicopter drop to an otherwise inaccessible location. In the latter example, the funding for transport and the availability of knowledgeable personnel willing to take the physical risk(s) of being "dropped" in a extreme wilderness situation can deplete a limited budget in a relatively short time. Helicopters may cost up to hundreds of dollars/hour.

Another ground truthing method is to use additional aerial photos or images not used in the original project to validate the results.

Peer Review

The cost of peer review can vary. The cost is dependent on the review being internal or external. Another factor affecting the cost of peer review is the extent to which the product is reviewed. The extent can range from soundness-of-conclusion to a "blind" evaluation. These are further defined in the next section.

Forms of Peer-Review

Peer Review can be very resource-intensive. Therefore, appropriate form of peer review should be considered for the category of process or product. Peer-review may take two forms:

Soundness of Conclusion

In this type of peer review a scientists' conclusions are evaluated based on a review of the written report.

Blind Evaluation

In this type of peer review, another scientist is given the same data and metadata and asked to produce their own photo interpretation. The two conclusions are compared.

In addition there are internal and external reviews. Broadly speaking, these reviewers may have the same interests and work, essentially, from the same funding.

Therefore, internal reviewers may not be considered as "independent" "arms'-length" reviewers." Subclassifications of internal and external reviews are:

Internal -

1. Within an office - Shoulder-to-shoulder scientists - in this case a scientist, in the same office/laboratory environment, may conduct a soundness of conclusion review or a blind evaluation
2. Within an organization - Organizationally-ties personnel may include current contractors, cooperators, grantees. These personnel have the same interests and goals in common and may conduct a soundness of conclusion review or a blind evaluation

External -

1. Journal reviews - these often are solely "soundness of conclusion" reviews. The peer review policies, practices, and guidelines of each journal may vary greatly. In addition, the level of commitment of the individual reviewer may vary greatly. Therefore, these reviews may or may not be of the highest quality.
2. Expert (Big-name) reviews - these types of reviews are rarely pro bono. Expert reviewers may be associated with a University, corporation, or be independent consultants. Costs will vary for "soundness of conclusion" review or a "blind evaluation."
3. University, Corporation, Independent Consultant - similar to Experts, these reviewers operate independent from the matter at hand. Here again costs will vary for "soundness of conclusion" review or a "blind evaluation."

Merging Peer Review and QA Categories

The EPA "graded approach" substantiates the development of "categories" described above. Categories have been delineated and accepted throughout most of EPA. Here we combine the currently accepted tools of assessment, ground-truthing and peer review, and the different forms of peer review as applied to photo interpretation, and organize them into the categories used by the QA Community:

Category 1 - Internal (EPA) & external peer review combined with ground-truthing
Category 2 - Internal (EPA) & external (journal) peer review combined with ground-truthing
Category 3 - Internal peer review combined with ground-truthing
Category 4 - Internal peer review and/or ground truthing.

The above-listed categories present different kinds, or levels, of peer review. These categories may be extended and made available for managerial use.

The Utility of Peer Review/QA Categories

On April 25, 2003, the President authorized a new national policy that establishes guidance and implementation actions for commercial remote sensing capabilities - *U.S. Commercial Remote Sensing Space Policy*. The policy directs the U.S. Government to:

- "rely to the maximum practical extent on U.S. commercial remote sensing capabilities for filling imagery and geospatial needs for military, intelligence, homeland security, and

civil users;"

- "Focus U.S. Government remote sensing space systems on meeting needs that cannot be effectively, affordably, and reliably satisfied by commercial providers [9]

In October 2003, the US EPA circulated an internal "Issue Paper." The issue paper raises impediments to the implementation of the policy within the US EPA. One issued raised states that there is...

> "Insufficient knowledge at both the staff and managerial levels about how to best acquire and utilize commercial satellite data products and services limits current usage"[10]

The Peer Review/QA Categories described above afford a solution for managers who need to use interpretation of remotely sensed images, but who may lack knowledge about remotely sensing technology and the photo interpretation process. By selecting the category based on the utility of the data, that is, how the manager will use the image and interpretation, the level of peer review required is also selected.

Conclusion

By using categories to define the type of product, and applying the appropriate level of peer review, the use of peer review of a QA tool is realized. In addition, having a process such as this defined, and making it a part of Aregular business@ for the Agency, the courts are more likely to accept the photo interpretation product as evidence without additional scrutiny (by the trier).

In litigation defense, the Agency has a process to which they can refer and produce documentation to verify, the quality review of photo interpretation products - thus establishing the veracity of results.

As a managerial tool, the categories provide for managers to simply select the category to describe the type of product, substantiating documentation, and the level of peer review required to meet their needs.

9*U.S. Commercial Remote Sensing Space Policy*, Office of Science and Technology Policy, The White House

10*Issue Paper: U.S. Commercial Remote Sensing Space Policy: Civil Agency Implementation Plan* MacWilliams, Lauren, US EPA Internal Report, October 10, 2003

Data Quality Objectives and Measurement Quality Objectives for Research Projects

Robert S. Wright, Mail Code E343-03
USEPA, National Risk Management Research Laboratory
Air Pollution Prevention and Control Division, Research Triangle Park, NC

Abstract: This paper provides assistance about systematic planning using measurement quality objectives (MQOs) to those working on research projects. These performance criteria are more familiar to researchers than data quality objectives (DQOs) because they are more closely associated with the measurement systems being used. Because of the diverse nature of research, it is not possible to describe cookbook-style procedures for developing these criteria. Instead, several general concepts and techniques are presented and researchers can choose those techniques that best fit their specific applications.

Introduction

Under the EPA Quality System, environmental programs are required to use a systematic planning approach to develop acceptance or performance criteria for the collection, evaluation or use of environmental data. EPA developed the DQO process to assist systematic planning. While not mandatory, this process is the recommended planning approach for many environmental data collection activities. It is based on the assumption that the ultimate goal for these activities is to make some decision (e.g., a regulatory compliance determination). The process uses a statistical approach to establish acceptance limits for decision errors and to develop a sampling and analysis plan to collect data with uncertainties within these limits.

It is often a challenge to apply the DQO process to the development of acceptance and performance criteria for research projects. In many cases, there is no decision to make, although there may be some requirements for the uncertainty of measurements. For basic research projects, the measurement system may not be developed enough for the uncertainty to have been characterized, even for individual components of the system. Other procedures may be needed to develop the required acceptance and performance criteria.

Distinctions between DQOs and MQOs

There is some confusion about DQOs and MQOs. Some people incorrectly regard these terms as being equivalent. Forthcoming EPA guidance on data quality indicators (EPA QA/G-5i) states that MQOs should be developed as an integral part of the sampling and analysis design generated during the final step of the DQO process. It makes the following distinction:

> "...MQOs are not synonymous with project DQOs. DQOs establish the full set of specifications for the design of the data collection effort. The design typically incorporates and specifies requirements for total variability. These requirements are used, in turn, to establish performance criteria, stated as MQOs, for significant components of total variability..."

For example, EPA has established a National Ambient Air Quality Standard for the maximum 8-hour mean concentration of ozone in the atmosphere. A reasonable DQO for this program might be that there can be no more than a 5-percent probability of making an incorrect decision (i.e., a false positive) based on ozone measurements that indicate an urban area has not attained the standard. In addition to the geographical and temporal variability of ozone concentrations in urban air, the decision makers must also consider the uncertainty of the ozone measurements. Appropriate MQOs for the measurements might be a bias of no more that 10 percent of the mean concentration and a precision of no more than 10 percent of the mean.

One way to distinguish between DQOs and MQOs is that the former are associated with data users and the latter are associated with data collectors. Another way is that DQOs function at the level of project goals, while MQOs function at the level of measurement system capabilities. A decision maker balances the risk of making an incorrect decision against the cost of the data that allow the decision to be made. The DQO process is a formal balancing mechanism that uses statistical techniques. The outcome of this systematic planning is DQOs that are consistent with project goals. A data collector balances the uncertainty of the measurements against the costs of the sample and analytical procedures that are used to collect the data.

A data collector should use some scientifically defensible process to link the desired uncertainty with the capabilities of specific sampling and analytical procedures that are used in the project. The first outcome of this process is the identification of data quality indicators (DQIs), such as bias, precision, and representativeness, which are quantitative or qualitative parameters that characterize the uncertainty of the project's measurement systems. The final outcome of the process is MQOs, which are specific goals for these DQIs. The MQOs are generally quantitative and must be verifiable by measurement during the project. If data will be collected from the environment, this process must also develop a sampling plan to select the number and location of the samples to be collected to attain some desired level of uncertainty due to the heterogeneous nature of the sample population.

Under ideal circumstances, the data users are identified and their data quality needs are determined. They participate in systematic planning by developing DQOs that are based on project goals. Data collectors design measurement system MQOs and estimate funding that will be needed to attain the DQOs. The process balances the needs and capabilities of both groups. Several iterations may occur before a mutually acceptable set of DQOs and MQOs is established.

For research projects, it may be difficult to identify the data users or to establish a desirable level of uncertainty for the data. It may be necessary to develop DQOs and MQOs that are based only on the capabilities of the measurement systems. Nevertheless, researchers need to have QC procedures that allow them to verify that the measurement systems are operating correctly and to be able to estimate the uncertainty of the environmental data that they collect.

Accounting for Natural Sample Variability in Environmental Measurements

For the purposes of this discussion, it will be assumed that the measurements to be made do not involve collecting samples from the environment and that the uncertainty associated with spatially or temporally heterogeneous samples does not have to be considered in the systematic planning. In reality, the uncertainty associated with sample variability must be considered, as

well as that associated with the measurement systems, in the systematic planning for a project and the development of its DQOs.

If the sample variability is large relative to the measurement variability, efforts to reduce the magnitude of the measurement uncertainty may not be cost-effective. It may be more fruitful in this instance to collect a larger number of more uncertain measurements than a smaller number of less uncertain measurements. Conversely, if measurement variability is large relative to sample variability, then efforts to reduce measurement uncertainty may be appropriate.

DQOs and MQOs need to be Realistic, Measurable, and Auditable

QA project plans should contain DQOs and MQOs that represent realistic data quality needs and measurement capabilities for the project. Generic DQOs and MQOs whose attainment cannot be verified during the project should not be included in a plan because they serve no purpose. Values for MQOs that are taken from the technical literature should be accompanied by information about the conditions under which the MQOs can be considered to be applicable. The project staff's hopes or unsubstantiated guesses regarding data quality are not adequate bases for DQOs and MQOs. The following table presents MQOs for a hypothetical project.

Measurement Parameter	Analysis Method	MQO for Bias	MQO for Precision	MQO for Completeness
Parameter A	Method A	+/- 5 %	+/- 10 %	+/- 90 %
Parameter B	Method B	+/- 20 %	+/- 20 %	+/- 90 %

These MQOs may or may not be realistic. One cannot tell from the table alone. The QA project plan needs to present mathematical formulas that define how project staff will determine by measurement whether the MQOs have been attained. The plan also needs to present specific QC check procedures that will be used to determine MQO attainment as well as the specific acceptance criteria for these procedures. It does no good to establish a MQO for bias if there are no credible reference standards available for checking the bias of the method. If the plan describes such QC check procedures, acceptance criteria, and reference standards, then project staff will be able to determine whether data quality is adequate for the intended use of the data. Additionally, an internal or external auditor can verify independently whether MQOs were attained by performance evaluations or by review of QC check results obtained by project staff.

There is a feedback loop between the establishment of DQOs and MQOs and the reconciliation of measurement data with these quality objectives. Both parts of the process are necessary to establish the quality of the data for the project. Doing this reconciliation while data are being collected can lead to improvements in the measurement systems. Additionally, the reconciliation for one project can help the development of DQOs for a follow-on project.

DQOs may be stated in quantitative or qualitative terms. Generally, quantitative statements are preferable, but acceptable qualitative DQOs are possible, such as the following:

"The project intends to produce data that will qualify to receive the 'A' rating with respect to

the rating system described in Section 4.4.2 of the Procedures for Preparing Emission Factor Documents (EPA-454/R-95-015)".

Although it is stated in qualitative terms, this DQO is measurable using specific acceptance criteria that are presented in the cited document. Because these criteria include whether EPA test methods were used for the measurements, reasonable MQOs for the project may include the QC check acceptance criteria specified in the methods. There is a direct link between quantitative MQOs and the qualitative DQO. At the end of this project, data collectors and data users (e.g., stakeholders, regulators, and management) can determine whether the data quality is acceptable. Contrast this qualitative DQO with the following qualitative DQO, which is not acceptable:

"For this project, the qualitative data quality objective is to provide data to assess emissions related to the operations of the source. This QA project plan is a product of a systematic planning process and it contains the information needed to carry out the field operations and measurements in order to meet this DQO."

There is no way to tell whether the resulting data are suitable for the intended use because acceptable criteria have not been defined and a procedure for assessing data quality has not been developed. The best thing that project staff can do after data have been collected is to carefully characterize the uncertainty, but this process cannot be considered to be systematic planning.

Instrument Performance Specifications as MQOs

Lacking other information, project staff may wish to use instrument performance specifications taken from sales literature or operating manuals as MQOs. Such specifications must be viewed with some skepticism because they may not have been determined objectively and rigorously. A vendor may feel compelled to present an instrument's performance in the best possible light relative to that of instruments offered by competitors. Various uncertainty components may have been omitted from the specification or the specification may be based on measurements obtained under tightly controlled laboratory conditions not typical of routine service (e.g., daily calibrations of an instrument normally calibrated on a weekly basis). It is preferable that instrument performance be evaluated using objective and written procedures that are widely accepted for instruments of that type. The measurement of performance by an independent, objective evaluator is generally regarded as credible evidence.

If the instrumentation vendor can provide credible evidence about how the performance specification was determined and about the measurements that were used in the determination, the use of such specifications as MQOs is reasonable. In these cases, the instrument's operating manual may yield QC check procedures and acceptance criteria that can be used directly by project staff. If these procedures are followed during the project and if the criteria are attained, one may conclude that the data uncertainty corresponds to the instrument's specifications.

Error Analysis

For well-characterized measurement systems, project staff should be able to develop a functional relationship between the MQOs for a measurement system and the DQOs for the project. Error propagation techniques allow the uncertainties of individual measurement system components (expressed as standard deviations) to be combined into an estimate of the overall measurement uncertainty. This approach assumes that the major sources of measurement variability have been identified, that bias can be controlled, and that relevant QC checks can be developed to characterize the variability. If a relationship exists, then one can demonstrate that DQOs have been attained if all QC check results fall within the corresponding MQOs.

Consult the bibliography for detailed information about measurement uncertainty and error propagation calculations. The basic error propagation formulas have implicit assumptions such as the independence of the measurements, their randomness, and their variances being small. Significant deviations from these assumptions will lead to significant errors in the uncertainty estimates made using these formulas. When in doubt about the use of statistical calculations in specific applications, it's always a good idea to consult with a statistician.

Limitations of Error Propagation Techniques

Uncertainty estimates and MQOs derived from propagation of error calculations are useful as long as one can determine the major sources of measurement error and can quantify the magnitude of the uncertainty associated with each source. In a research project, the measurement system may be so new that its performance is not yet characterized. In these circumstances, empirical approaches may be needed to determine measurement uncertainty and to establish MQOs for a project. For instance, collocated measurements can characterize the precision of measurement systems. Performance evaluations can characterize the bias of the systems. The results of such method evaluations can be used to establish defensible MQOs for future projects, even in the absence of knowledge of error sources and uncertainty magnitudes.

Error propagation calculations cannot be used to combine individual QC check results to yield an uncertainty estimate. One can't substitute a single measured value for a standard deviation. The best that can be done is to use these results to demonstrate that MQOs have been attained and to conclude that the uncertainty is within the acceptance criterion. The means and standard deviations for multiple QC checks can be combined to yield an uncertainty estimate.

Guide to the Expression of Uncertainty in Measurements

In recent years, the international metrology community has standardized the methods for calculating uncertainty in the *Guide to the Expression of Uncertainty in Measurement*, commonly referred to as GUM. In the United States, standards bodies such as American National Standards Institute (ANSI), National Conference of Standards Laboratories (NCSL), and National Institute of Standards and Technology (NIST) have adopted GUM as their official method for calculating uncertainty for metrology in testing and calibration laboratories. The statistical techniques in GUM provide a standard basis for determining uncertainty in research projects. The statistical terminology that is used in GUM differs somewhat from customary statistical terminology.

Uncertainty Budgets

One can use error propagation calculations to develop an uncertainty budget for the measurement system. Given a particular requirement for overall uncertainty, each component of the measurement system can be allocated an appropriate portion of this uncertainty. Such an approach allows project staff to design measurement systems with MQOs that will allow attainment of DQOs or to assess whether it is technically feasible to attain the proposed DQOs. It also allows project staff to develop the most cost-effective method to reduce the overall uncertainty. They can determine which measurement system component has the largest effect on the overall uncertainty and then concentrate their efforts on improving the quality of the data from that component. They can also use uncertainty budgets to evaluate alternative measurement strategies and choose the most cost effective strategy. For example, it may be less expensive to attain a DQO by making multiple measurements with a cheap, low precision method than to do so making a single measurement with a expensive, high precision method.

Bibliography

American National Standards Institute and National Conference of Standards Laboratories, 1997. *American National Standard for Expressing Uncertainty- U. S. Guide to the Expression of Uncertainty in Measurement.* ANSI-NCSL Z540-2-1997, Second Edition, National Conference of Standards Laboratories, Boulder, CO.

Coleman, H.W. and Steele, W.G., 1999. *Experimentation and Uncertainty Analysis for Engineers*, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ.

Kimothi, S.K., 2002. *Uncertainty of Measurements: Physical and Chemical Metrology: Impact & Analysis*, American Society for Quality, Milwaukee, WI.

National Institute of Standards and Technology, 2004. *NIST/SEMATECH e-Handbook of Statistical Methods,* (see http://www.nist.gov/stat.handbook/)

Taylor, B.N. and Kuyatt, C.E., 1994. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297, Gaithersburg, MD. (see http://physics.nist.gov/cuu/Uncertainty/)

Taylor, J.R., 1997. *An Introduction to Error Analysis. The Study of Uncertainties in Physical Measurements*, Second Edition, University Science Books, Sausalito, CA.

Greater Consistency of Radiation Emergency Response Analytical Methods Using a Performance Based Approach

Ben Hull, U.S. EPA Office of Radiation and Indoor Air
John Griggs, U.S. EPA, Office of Radiation and Indoor Air, NAREL

The paper will focus on the development of radiation Emergency Response Measurement Quality Objectives (MQOs) and how these Measurement Quality Objectives can be incorporated into Emergency Response planning documents to promote comparability of data. The process for the development of Emergency Response MQOs is based on the Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP) Manual which provides guidance for the planning, implementation and assessment phases of those projects which require the laboratory analysis of radionuclides. Measurement Quality Objectives (MQOs) are statements of performance objectives or requirements for a particular analytical method performance characteristic. MQOs can be viewed as the analytical portion of the overall project DQOs. In a performance based approach, the MQOs are used initially for the selection and evaluation of analytical protocols and are subsequently used for the ongoing and final evaluation of the analytical data. In MARLAP, the development of MQOs for a project depends on the selection of an action level and gray region for each analyte during the directed planning process. For emergency response, planning different decisions include whether an individual should evacuate or shelter, whether the water is safe to drink, and whether land can be released for unrestricted use. For each of these decisions, action levels are used to determine measurement quality objectives which are then used to develop analytical schemes and laboratory protocols.

Introduction

The Environmental Protection Agency, Office of Air and Radiation's Radiological Emergency Response Team (RERT) is required under the National Oil and Hazardous Substances Pollution Contingency Plan (NCP) to provide mobile and fixed laboratory capabilities and train personnel for technical support in response to radiological/nuclear incidents that results in radiation exposure to the public and the environment. While the fixed analytical capability at Office of Air and Radiation's (ORIA's) two laboratories is a critical part of the Agency's response capability, these laboratories would be quickly overwhelmed in the event of a major radiological terrorist event, or should multiple events occur at the same time. Because of this, use of private, federal, and state radioanalytic capabilities will be required to support the analysis of the potential tens of thousands of samples collected for the purpose of monitoring and assessment during a major incident.

To handle these analysis needs and the associated data quality requirements, a core network of laboratories is needed that will provide consistent and comparable data of the appropriate and required quality. This network will provide necessary reliable data to Federal, state, and local responders during an incident. In order to develop this network the following activities need to be accomplished: Development of Data Quality Objectives (DQOs) and Measurement Quality Objectives (MQOs) for emergency response radioanalytical methods; Assessment of the

capability and capacity of available private sector radiation laboratories relative to emergency response; Assessment of radiation laboratories relative to the emergency response radioanalytical requirements; Development and assessment of performance evaluation and audit requirements specific to emergency response; Development of electronic data reporting capability consistent with the Agency's organic and inorganic analysis capability.

In addition having well defined measurement quality objectives and emergency response radioanalytical methods will directly support the objective of the Federal Radiological Emergency Response Plan (FRERP) which is to establish an organized and integrated capability for timely, coordinated response by Federal agencies to peacetime radiological emergencies. Under the FRERP, EPA is the responsible agency to coordinate Federal offsite radiological monitoring and assessment for the intermediate and the long term response. The existence of a coordinated federal and state emergency response infrastructure to address data analysis and information management will also address Homeland Security Presidential Directive 9 (HSPD-9) which establishes a national policy to defend the agriculture and food system against terrorist attacks, major disasters, and other emergencies. Specifically tasks within HSPD-9 include the development of coordinated surveillance and monitoring systems for water quality and the development of a nationwide laboratory network for water quality that integrates Federal and State laboratory resources.

MARLAP Overview

The Environmental Protection Agency (EPA) is finalizing the development of the Multi-Agency Radiation Laboratory Protocols Manual (MARLAP) and a Radionuclides Inventory Manual. MARLAP and the Radionuclides Inventory will support a number of EPA programs and will provide guidance in all the major areas of radioanalytical laboratory work and project planning including emergency response.

The Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP) Manual is a document which provides guidance for the planning, implementation and assessment phases of those projects which require the laboratory analysis of radionuclides. MARLAP's basic goal is to provide guidance and a framework for project planners, managers and laboratory personnel to ensure that radioanalytical laboratory data will meet a project's or program's data requirements and needs. To attain this goal, the manual seeks to provide the necessary guidance for national consistency in radioanalytical work in the form of a performance-based approach for meeting a project's data requirements. The guidance in MARLAP is designed to help ensure the generation of radioanalytical data of known quality appropriate for its intended use.

MARLAP was developed by a workgroup which included representatives from the Environmental Protection Agency (EPA), the Department of Energy (DOE), the Department of Defense (DOD), the Nuclear Regulatory Commission (NRC), the National Institute of Standards and Technology (NIST), the U.S. Geological Survey (USGS), the U.S. Food and Drug Administration (FDA). State participation in the development of the manual involved contributions from representatives from the Commonwealth of Kentucky and the State of

California. The draft manual is currently available at the following website: http://www.eml.doe.gov/marlap/.

Performance-Based Approach

MARLAP provides the necessary guidance for using a performance-based approach to meet a project's analytical data requirements. In a performance-based approach, the project-specific analytical data requirements that are determined during directed planning serve as measurement performance criteria for analytical selections and decisions. The project-specific analytical data requirements also are used for the initial, ongoing, and final evaluation of the laboratory's performance and the laboratory's data. MARLAP provides guidance for using a performance-based approach for all three phases, planning, implementation and assessment, of the data life cycle for those projects that require radioanalytical laboratory data. This involves not only using a performance-based approach for selecting an analytical protocol, but also using a performance-based approach for other project activities, such as developing acceptance criteria for laboratory quality control samples, laboratory evaluations, data verification, data validation, and data quality assessment.

There are three major steps or processes associated with a performance-based approach. The first is clearly and accurately defining the analytical data requirements for the project. The second involves using an organized, interactive process for selecting or developing analytical protocols to meet the specified analytical data requirements and for demonstrating the protocol's ability to meet the analytical data requirements. The last major activity involves using the analytical data requirements as measurement performance criteria for the ongoing and final evaluation of the laboratory data, which would include data verification, data validation, and data quality assessment. MARLAP provides guidance in all three of these areas. Within the constraints of other factors, such as cost, a performance-based approach allows for the use of any analytical protocol that meets the project's analytical data requirements. For all relevant project activities, the common theme of a performance-based approach is the use of project-specific analytical data requirements that are developed during project planning and serve as measurement performance criteria for selections, evaluations, and decision-making.

Performance Objectives: Data Quality Objectives and Measurement Quality Objectives

One of the outputs of a directed planning process is DQOs for a project or program. DQOs are qualitative and quantitative statements that clarify the study objectives; define the most appropriate type of data to collect; determine the most appropriate conditions from which to collect the data; and specify tolerable limits on decision error rates (ASTM D5792; EPA, 2000). DQOs apply to all data collection activities associated with a project or program, including sampling and analysis. In particular, DQOs should encompass the "total uncertainty" resulting from all data collection activities, including analytical and sampling activities.

From an analytical perspective, a process of developing the analytical data requirements from the DQOs of a project is essential. These analytical data requirements serve as measurement performance criteria or objectives of the analytical process. MARLAP refers to these performance objectives as "measurement quality objectives" (MQOs). The MARLAP Manual provides guidance on developing the MQOs from the overall project DQOs. MQOs can be viewed as the analytical portion of the DQOs and are therefore project-specific. MARLAP provides guidance on developing MQOs during project planning for select method performance characteristics, such as method uncertainty at a specified concentration; detection capability; quantification capability; specificity, or the capability of the method to measure the analyte of concern in the presence of interferences; range; ruggedness, etc.

An MQO is a statement of a performance objective or requirement for a particular method performance characteristic. Like DQOs, MQOs can be quantitative and qualitative statements. An example of a quantitative MQO would be a statement of a required method uncertainty at a specified radionuclide concentration, such as the action level—i.e., "a method uncertainty of 3.7 Bq/kg (0.10 pCi/g) or less is required at the action level of 37 Bq/kg (1.0 pCi/g)." An example of a qualitative MQO would be a statement of the required specificity of the analytical protocol—the ability to analyze for the radionuclide of concern given the presence of interferences—i.e., "the protocol must be able to quantify the amount of ^{226}Ra present given high levels of ^{235}U in the samples." The MQOs serve as measurement performance criteria for the selection or development of analytical protocols and for the initial evaluation of the analytical protocols. Once the analytical protocols have been selected and evaluated, the MQOs serve as criteria for the ongoing and final evaluation of the laboratory data, including data verification, data validation, and data quality assessment. In a performance-based approach, analytical protocols are either selected or rejected for a particular project, to a large measure, based on their ability or inability to achieve the stated MQOs. Once selected, the performance of the analytical protocols is evaluated using the project-specific MQOs.

Emergency Response

Based on combining potential emergency response scenarios and associated radionuclides of interest with different decisions that need to be made for each scenario, one can develop a range of data quality objectives and measurement quality objectives for emergency response radioanalytical methods. Potential release scenarios are varied but examples include an alpha radiation dispersal device scenario or a beta/gamma radiological dispersal device which could potentially contaminate a few square miles of land, and a water security scenario which includes disruption of water and wastewater systems and contamination of tens of thousands of gallons of water, potentially requiring the analysis of 1000s of samples.

The estimated analytical needs for the radiological dispersal device (dirty bomb) scenario reveal that 15,000 to 20,000 analyses will be required over a very short time frame (a couple weeks to a month) and would include a combination of rapid screening analyses followed by nuclide specific analyses on high priority samples. In order to meet these high analytical requirements multiple laboratories would be needed to produce data of comparable quality.

Project-specific analytical data requirements would be used to ensure that each laboratory is producing data of known and appropriate quality. These requirements would be developed during project planning and serve as measurement performance criteria for selecting or developing analytical protocols to meet the specified analytical data requirements, for demonstrating the protocol's ability to meet the analytical data requirements, and the ongoing and final evaluation of the laboratory data.

Based on the emergency response decisions which include whether an individual should evacuate or shelter, whether land can be released for unrestricted use, whether the water is safe to drink, or whether the water is safe to use different DQOs will be developed. For each of these decisions, action levels are used to determine measurement quality objectives which are then used to develop analytical schemes and laboratory protocols.

An example of a quantitative MQO would be a statement of a required method uncertainty at a specified radionuclide concentration. For the example scenario with a radiological dispersal device with Sr-90 contamination, different action levels for Sr-90 may range from an evacuation decision action level at 100 pCi/g to an action level for unrestricted use at approximately 1 pCi/g. For the first case a method uncertainty of 10 pCi/g or less is required at the action level of 100 pCi/g; and for the second case a method uncertainty of 0.10 pCi/g or less is required at the action level of 1.0 pCi/g.

Conclusion

The EPA Lessons Learned Report identified laboratory infrastructure and data analysis as critical needs for the Agency. Specifically the report identified the following needs: consistent data from multiple agencies, greater consistency of analytical and quality assurance methods; increasing national laboratory analysis capacity, and decreasing laboratory response time. In order to address these issues for radiologically analyses, a radiation response laboratory network combined with a performance-based approach for radioanalytical data requirements is needed.

The performance-based approach for analytical data requirements will meet a range of radionuclide release scenarios and potential decisions that need to be addressed. The MARLAP Manual provides the foundation for the development of this performance-based approach. The radiation response network composed of multiple laboratories will address the need of the extensive analytical requirements and the critical time commitments. The response network should also be of sufficient size and capacity to address the range of potential emergency response scenarios so that performance measures for capacity and capability are met. Performance measures for capability and capacity would include: the percentage of samples which the surge capacity laboratories can analyze in the specified time requirements based on the potential response scenario; and the percentage of surge capacity laboratories that meet EPA Emergency Response laboratory competency requirements. In addition, there is the continuing need to create awareness, foster understanding, and promote the use of Emergency Response measurement quality objectives and analytical protocol specifications with Federal, State, Local government and private laboratories.

References

MARLAP, August 2001. *Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP) Manual*, *Draft for Public Comment*. EPA 402-B-01-003, NUREG-1576.

U.S. Environmental Protection Agency, February 2002. *Lessons Learned in the Aftermath of September 11, 2001.*

Ben Hull, U.S. EPA Office of Radiation and Indoor Air (6608J), 1200 Pennsylvania Ave., Washington, District of Columbia, 20460, Email: hull.benjamin@epa.gov

John Griggs, U.S. EPA, Office of Radiation and Indoor Air, NAREL, 540 South Morris Avenue, Montgomery, AL 36115-2601, Phone Number: (334)270-3450, Email: griggs.john@epamail.epa.gov

A CD-ROM Based Quality Assurance Project Plan Preparation Tool for Tribes

David R. Taylor and Vance S. Fong
U.S. EPA Region 9 Quality Assurance Office, San Francisco, California
Contact David Taylor at 415-972-3803 or Taylor.David@epa.gov

Abstract:

A Work Group has been working on developing a tool that could be used by tribes to assist them in preparing Quality Assurance Project Plans for EPA. This project is the result of a number of concerns expressed by tribes with EPA's approach to Quality Assurance. Among these are:

1. *Tribes value QAPPs and the need to have good quality, defensible data, but view the current QAPP process as too labor and resource intensive.*

2. *Quality Assurance Project Plan requirements are not standardized throughout the Agency. Tribes find that QAPP formats are inconsistent among the different Regions and within some EPA Regions.*

3. *Lack of resources provided to assist Tribes with QAPP development and implementation are generally inadequate. It is currently very costly (labor and resource intensive) for Tribes to develop and implement QAPPs. As a result, Tribes often spend much needed funds developing several iterations of a QAPP with little assistance or direction (e.g., Regional assistance in identifying QAPP requirements and assisting with development, including where changes need to be made to produce an "adequate" plan).*

To address these concerns a Work Group was formed which has been working to develop CD-ROM based approach to QAPP development. The Work Group consists of representatives from EPA Regions, tribes, and organizations which work with tribes. The initial focus has been on water monitoring projects, but the long range goal to is make the tool usable for other programs and media under which tribes receive grants. If successful, the program it is expected that the CD-ROM approach could be easily changed to accommodate other small grantees, which often have problems similar to tribes. This session will provide a up-date on the current development effort, and it is planned to have a preliminary version to demonstrate.

Electronic Research Notebooks

Shirley J. Wasson, Mail Code E343-03
USEPA, National Risk Management Research Laboratory
Air Pollution Prevention and Control Division, Research Triangle Park, NC
and
Brenda Culpepper, Mail Code B343-01
USEPA, National Health and Environmental Effects Research Laboratory
National Accountability and Research Management Staff, Research Triangle Park, NC

Research notebooks contain intellectual property which must be guarded until it can be disseminated with credit attributed to its creators. Electronic notebooks (EN) have advantages over paper. Advances in computing in the last decade mean that the entire research and invention process is executed in electronic files beginning with the plans (word processor), continuing with the implementation (instrument output files, spreadsheets, image files, mathematical calculations), and ending with the reports (visualization programs, reports, presentations). Paper notebooks create an added burden by requiring that much of the electronic record be printed and cut and pasted into the notebook.

Paper notebooks are easily secured because just one record exists, while EN potentially exist in many versions. Their usability as research notebooks relies on the ability to integrate into the records a computer-generated, automatic time-stamped audit trail and transparent access control. These last ten years have produced several commercial products that seek to provide irrefutable data integrity and security

EPA has proposed rules for electronic record keeping via the Cross-Media Electronic Reporting and Record-keeping Rule (CROMERRR) and has initiated several demonstration projects in cooperation with the regulated community. The Agency has not taken up the issue of electronic research notebooks for either personal or collaborative use.

This paper details the public availability of EN and an example of a system in use within a research laboratory in the Office of Research and Development.

Introduction

The advent of the affordable personal computer, networks, and the searchable world wide web has brought the possibility of replacing paper research notebooks with electronic notebooks (EN). Bound paper research notebooks are kept by scientists, engineers, and technicians as written records of their work.Additionally, paper notebooks are used for sample logs and instrument maintenance log books. Their versatility has been sufficient for their intended purpose for several hundred years, and until the advent of modern computers were the state of the art. They are repositories of intellectual property, carefully guarded until published or patented and credited to its creator. By nature they are generally inaccessible to remote research collaborators. They are unique, secure, and limited.

Definition

The Collaborative Electronic Notebook Systems Association (CENSA) is an international industry association that seeks to improve automated technologies and methods that facilitate scientific collaboration among both technical and business people. CENSA provides a consensus definition of EN.

> **An electronic notebook is a system to create, store, retrieve, and share fully electronic records in ways that meet all legal, regulatory, technical, and scientific requirements.**

The definition emphasizes that the EN system consists not only of cutting edge technologies but a workable combination of policies, procedures, and regulations necessary to secure intellectual property. Since many of these either do not exist or are inconsistent from facility to facility and country to country, implementation of an EN system can be challenging.

Scope of Electronic Notebooks

One of the most useful attributes of EN is that they can be shared. The amount of intended sharing has a significant role in determining their complexity. A personal EN is kept primarily for the reasons that a personal bound written notebook is kept, to record progress on private, non-collaborative research, to log samples, or to log maintenance on an instrument. The notebook resides on one computer accessed only by the creator. Any file sharing is performed solely at his discretion. The collaborative notebook, on the other hand, is created to be shared among colleagues working on the same project or performing similar kinds of research, each contributing, reviewing, communicating, and benefitting from the project. It will reside on a server that may be accessed and manipulated by any authorized person either through a network or through a web browser. Once the information is uploaded to the collaboratory, however, it is out of the personal realm. It must be secured against premature disclosure and unauthorized manipulation. The notebook of record resides on one computer/server and is licensed. Only authorized personnel with controlled access have write capabilities on that computer. It may have a wider audience in other interested persons who have read-only capabilities.

Paper versus Electronic

Because EN are much more encompassing than a paper notebook, they require software systems that can provide access to and manipulate large amounts of data. Data are fed from source systems such as handheld field instruments, data acquisition systems, and laboratory analytical systems into data warehouses. Data from the warehouses are sorted, reformatted, and placed into databases which can be very large. Good EN systems will isolate the database maintenance functions from other areas of the notebook and keep those functions transparent, but automatic, such as automatic labeling and indexing. The notebooks function through interfaces. For instance, literature searches are easily conducted and stored using their web browsing and searching capabilities. They can retrieve, manipulate, and store references from online libraries,

databases, and web pages. The information is available in a form to create records, manage data, share files, and easily converted to provide planning documents, reports, and publications. The most labor intensive tasks performed manually are data entry, manipulation, and reporting. Replacing the manual-intensive manipulation of paper records with more costly electronic systems can result in a significant return on investment if the productivity gains in functionality and versatility are not outweighed by the difficulty of the learning curve.

The limitations of EN generally involve data integrity, i.e. the data warehouse containing incomplete/incorrect/incomprehensible/inconsistent records, missing records, missing fields, etc. or incorrectly extracting, transforming, or loading it from the source systems. The software may aggregate or calculate the data incorrectly. Users also have limited ability to view documents from electronic media simultaneously for visual comparison purposes, a disadvantage at audit. The software needs to have the ability to call up different types of data and view them via split screens and side-by-side monitors.

Functional Requirements for Electronic Notebooks

EN meet the National Archives and Records Administration (NARA) definition of a record.

> **Records include all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law, or in connection with the transaction of public business, and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government, or because of the informational value of the data in them.**

The rules for federal records drive the functional requirements for EN for governmental organizations.

Archival--If the research project and the principal researcher are based at EPA, the record will be the EN based at EPA, and will be archived on the EPA records retention and disposition schedules, 501, 502, 503, and 507.

Authentic–The principal researcher authenticates the record by ensuring that it is accurate, complete, and appropriate. Alteration or change to a previous entry must only be made as a new entry stamped with the current date and cross-referenced to the previous entry. Each entry must be signed and dated by the person who made the entry, and also by any others who contributed. If an invention is recorded, the true inventors must write “invented by” before their names, and the notebook reviewed and witnessed.

Retrievable--The notebook must be retrievable and usable during its life cycle. Authorized users are able to retrieve desired data or documents using an indexing or text search system based on key fields. Software is retained and lasting media are used.

Usable–The notebook must be stored on appropriate media and systems retained in a usable format until final disposition or destruction.

Secure–Only authorized persons have access. The system secured by nightly backup to network or CD, secured by password and lock and key and recoverable. Risk of unauthorized alterations or erasures is minimal.

EPA's Proposed Rule

EPA has proposed electronic record regulations via the Cross-Media Electronic Reporting and Record-keeping Rule (CROMERRR). It would provide the legal framework for electronic reporting and record-keeping under many of EPA's environmental regulations and would also set standards for electronic record-keeping under any EPA program. EPA is already receiving electronic record data via the Central Data Exchange (CDX). Through CDX the Agency is integrating existing data collection and information systems from state and local governments using state-of-the-art technologies.

Although the focus of CROMERRR for electronic record-keeping is the regulated community, any electronic record including an electronic research notebook may eventually be required to conform. To date, however, the Agency has not issued policy for electronic research laboratory notebooks for either personal or collaborative use.

Data Standards

Electronic research notebooks need a research data taxonomy, standards for data entry for all types of relevant data, and data entry rules including what meta-data (data attributes, data quality indicators) will be recorded. CENSA and other entities attempt to facilitate consensus data standards among their constituency via presentations, lectures, and training courses. EPA through the Environmental Data Standards Council (EDSC) is addressing data standards in the context of its CROMERRR regulation. According to the EDSC a data standard is a documented agreement among organizations that share or exchange data, including representation, formats, and definitions. It provides a common vocabulary that contains data elements, data element definitions, and formats to be used by EPA and the regulated community. At the time of this writing, several data standards are out for review. Almost certainly data standards applicable to EPA research data will arise from the final promulgated standards from the EDSC.

Public Domain E-Notebook: A DOE Project

In 1996, three Department of Energy (DOE) laboratories scattered around the country initiated the DOE2000 Electronic Notebook project. Oak Ridge National Laboratory (ORNL), Lawrence Berkeley National Laboratory (LBNL), and Pacific Northwest National Laboratory (PNNL) collaborated to develop a modular, extensible, notebook architecture and make it available to any interested group for trial. Their vision included providing capabilities for multimedia input, automatic logging of instrument data,

searchable entries, mining of large scientific reference databases, sharing files across a research group, remote access, security, and digital signatures. Early prototype version 1.10 was released February 1998. The newest update is Version 1.12 released January 2003. Currently there are over 300 groups around the country using the software and providing feedback.

The notebooks have Common Gateway Interface (CGI) scripts and Java applets to receive objects. They have a base set of editors (data creation and input tools) and data viewers (data display/visualization tools) for common data types–text, images, video, sketches/diagrams, equations, XY graphs, etc. They allow the use of standard or custom CGI-based security services for authorization, authentication, encryption, notarized time-stamping, digital signing and witnessing. Accessibility is by web browser. The notebook software is written in Perl5 and works on Win XP, 2000, Me, NT, 98, Mac, and Unix platforms. It requires only that a web server and Perl5 be installed on the machine where the notebook resides. A demonstration of the notebook is available at http://www.csm.ornl.gov/~geist/java/applets/enote/ and a copy of the notebook software is available from Al Geist, at gst@ornl.gov.

Extending this effort is another project called Scientific Annotation Middleware (SAM) which addresses the complexity resulting from collaborative, cross-disciplinary, computing-intensive research. SAM includes components and services that enable researchers, applications, and software agents to create metadata and annotations about data objects and the semantic relationships between them. SAM was created to address the limited capability of EN to show non-chronological relationships between entries, to support complex searches, and to interact with programs as well as people. SAM 1.0 was released in June, 2003, and SAM 2.0 is in process.

Commercial Packages

The last ten years have produced several commercial products. Some products worth exploring have been developed by CambridgeSoft, GynSys, New Information Paradigms (NIP), NuGenesis, Synthematix, ChemOffice Enterprise, and Tripos. The fact that few of them have emerged as being clearly superior can be attributed to the complexity of providing the required services. NuGenesis Technologies Corporation (Westborough, MA) has targeted the overwhelming data management needs of pharmaceutical and biotechnology research organizations. They can capture a wide variety of data types and formats issuing from the proprietary software of data sources like laboratory analysis instruments and make them available to project collaborators for analysis, synthesis, and summarization in reports, publications, and presentations. Their Scientific Data Management System (SDMS) software can:

- capture instrument data direct from the source for use in the EN
- archive raw data and summary reports with traceability to original raw data on safe media
- directly import and export varied data types from varied platforms (Windows, Unix, MAC)

- control access for security and privacy
- provide a full audit trail with electronic signatures and witnessing
- time and date stamp all data capture
- create a database catalog automatically using predetermined selection criteria

- search functions able to find text, even that embedded within graphics.

NuGenesis Application Control Manager (ACM) provides a strategy for system wide management of Microsoft Excel templates and spreadsheets. They can be:

- automatically captured, cataloged and securely stored
- located and viewed, reviewed, signed, and witnessed
- accessed by authorized users with differing data views and security levels
- audit trailed within the spreadsheet

EPA/ORD/NHEERL/MED Electronic Research Notebook

An aquatic bioassay project designed to monitor small fish exposure to disinfection by-products at the Mid Continent Ecology Division in Duluth, MN, is being recorded primarily in a Cold Fusion/Oracle database EN (Batterman et al, 2004). Most work is being performed in the laboratory. Beginning at the age of 21 days, fish are exposed to toxicants in 28 aquaria monitored daily for temperature, water flow, and visual observations. Periodically during the exposure period, the fish are sampled, barcoded, measured for weight and length, and digitally photographed. Barcoded histology slides are prepared from sampled sections of the fish, examined for pathology, and bioassay reports are issued.

The back end of the EN for the project is an Oracle relational database. The data in the modules of the database are linked based on the scannable ID barcode number common to each data entry in each module. Oracle needs front end tools to enter data and helper applications to display the data. Cold Fusion is a front end tool for this project, a program permitting access to the database from a web site. The key to searching for information in this project is the barcode number. Oracle is a robust database management system, enforcing referential integrity of all the data. For instance, it does not permit orphaned records or invalid codes.

Setup and maintenance of the Oracle 8.0 version database requires two administrators, an information technology (IT) professional as the Oracle Database Administrator and the project manager (PM) as the end user system administrator. All eight people associated with the project were set up with user names and passwords to have both read and write access to the database. The PM maintains values for all lookup tables to populate the drop-down selection lists. He may set data entry “protocols” for each measure, including specific upper and lower warning limits. The comment fields for visual observations are more format free. They can handle sketches or photographs made in another application and saved as image files

User name and time stamp are internally recorded on all new records. Once a new entry is saved, it can no longer be modified or deleted. It can only be marked erroneous and documented, then re-entered. There is no capability for electronic signatures or witnessing. The database provides printed barcode labels to identify samples. The database is incrementally backed up daily and fully backed up weekly. Archive is achievable by printing to compact disk.

To illustrate the search capabilities of the notebook, a user can access the Histology module through a web site, using the barcode ID to pull up a slide. The database will provide such information as the relevant fish, aquarium, conditions, measurements, comments from the visual observations of growth and morphology, and the bioassay reports. Images saved in the exposure log can be located, loaded, and viewed for study.

Some parts of the project such as the field data sheets are still recorded on paper and hand entered into spreadsheets and into the database. Project personnel would like future data to be electronically gathered in the field and automatically uploaded to the database. Bioassay reports are downloaded daily, printed, and pasted into notebooks as a permanent record. The (Cold Fusion/Oracle database) EN can provide detailed printed audit reports.

Summary

EN are available in the public domain and as commercial packages. Other unique combinations of software have been cobbled together to meet specific needs. They fulfill the need to keep large amounts of data organized, search capable, and shareable. When EN become sufficiently reliable and user-friendly, they will be preferred to paper notebooks.

References

Batterman, A., D. Lothenbach, and S. Erickson, personal communication, January 22, 2004.

Central Data Exchange (CDX), U.S. EPA, http://epa.gov/cdx.

CambridgeSoft's E-Notebook Enterprise, www.cambridgesoft.com/solutions.enotebook.cfm.

Environmental Information Management System (EIMS), U.S. EPA, http://oaspub.epa.gov/eims.

Geist, A,. J. Schwidder, D. Jung, and N. Nachtigal, *ORNL Electronic Notebook Project*, http://www.csm.ornl.gov/~geist/java/applets/enote/.

Helfrich, J.P. *A Process for Electronic Laboratory Notebooks*, NuGenesis Technologies Corporation, Westborough, MA, http://www.NuGenesis.com.

Helfrich, J. P. *Intellectual Property Data Management in the "Paperless" Lab*, NuGenesis Technologies Corporation, Westborough, MA, http://www.NuGenesis.com.

Lysakowski, Richard. *Comparing Paper and Electronic Laboratory Notebooks*, Collaborative Electronic Notebook Systems Association (CENSA). http://www.censa.org.

Mullen, Rick. *Learning to Share in the Lab*, Chemical and Engineering News, October 27, 2003, http://www.cen-online.org.

U.S. National Archives and Records Administration, www.archives.gov/index.html.

Improved Data Quality Systems Using the Florida DEP Automated Data Processing Tool (ADaPT) and Environmental Data Management System (EDMS)

Juan Manzano, South Florida Water Management District (SFWMD), 1480 Skees Road, West Palm Beach, FL, Delia Ivanoff , SFWMD, Thomas Frick, Florida Department of Environmental Protection (FDEP), 2600 Blair Stone Road, Tallahassee, Fl, Richard M. Amano, Laboratory Data Consultants (LDC), 7750 El Camino Real, Suite 2L, Carlsbad, CA and Julio Paredes, LDC, 112 Kings Way, Royal Palm Beach, FL

One of the challenges facing environmental professionals today is the collection and assessment of large amounts of analytical data. Assessment of data quality is essential as multi-million dollar decisions are made based on the analytical results. In addition, sharing and accessing of data across multiple data users is critical to environmental programs. Standardization of electronic deliverables allows for data from multiple data collectors to be compiled into a single database and made available to numerous data users or stakeholders on a project.

To deal with these challenges, the South Florida Water Management District uses support software modules to allow for expedited review and assessment of analytical data. The expedited review facilitates earlier decisions for corrective action with laboratories, assesses trends in data quality outliers, shortens turnaround time for data availability, and saves labor costs.

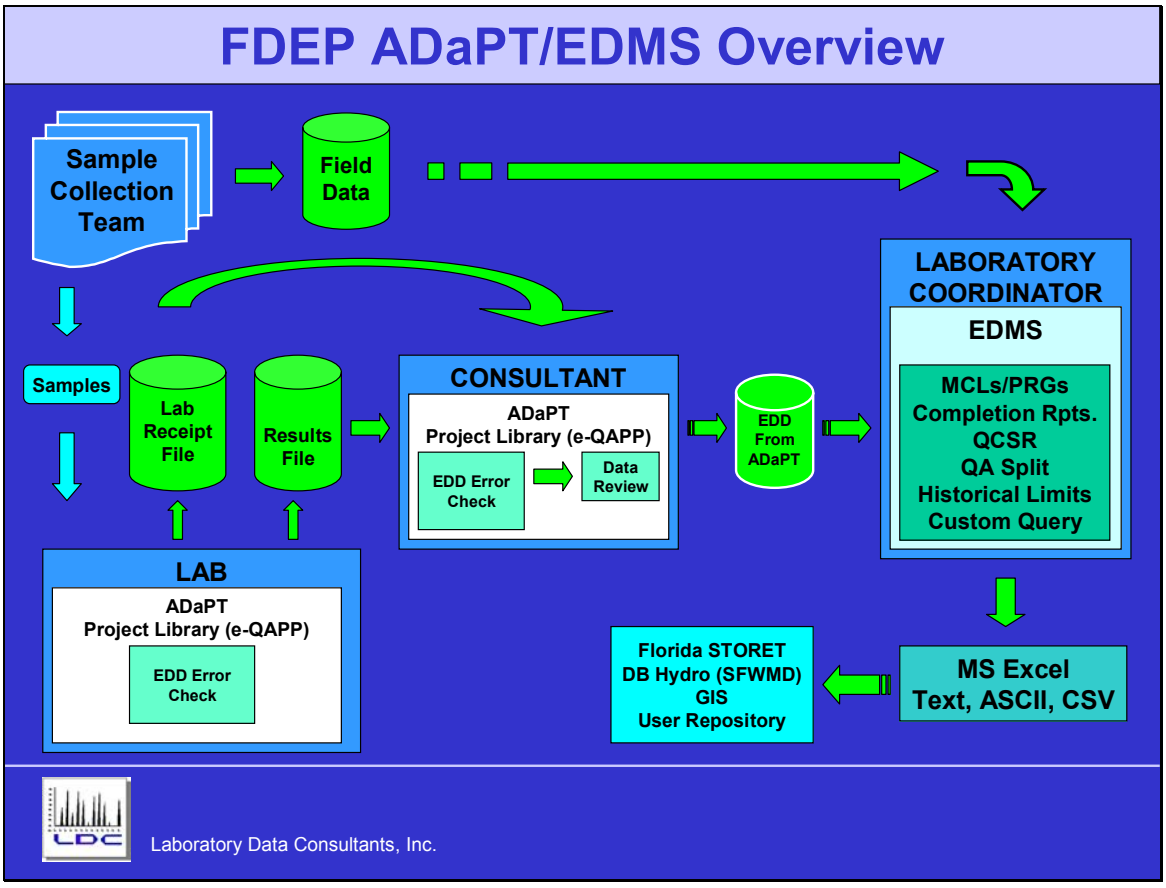
The support software consists of the Contract Compliance Screening (CCS), Automated Data Review (ADR), and Environmental Database Management System (EDMS) software programs developed by Laboratory Data Consultants, Inc. under contract to the Florida DEP. The software programs use an electronic data deliverable format based on data elements required for automated data review. The software was developed on a Microsoft ACCESS 2000/2002 platform. Customized modules perform automated data review equivalent to an EPA Level 3 evaluation and provide the user with discrete data qualification. The qualified data are exported into the EDMS module for further assessments such as historical data trending, and generation of Quality Control Summary Reports.

The ADaPT and EDMS software programs were developed as tools to support technical staff in the evaluation of analytical chemistry data using an expedited and cost effective automated process. A standardized format allows laboratories to streamline the data deliverable process. Using the CCS software, data deliverables can be verified immediately for completeness and compliance against project specific quality assurance criteria .The ADaPT and EDMS processing allows the data end user to efficiently evaluate large data sets for key indicators and ultimately determine the usability of the data.

Introduction

One of the major difficulties in consolidating, checking and comparing data from different laboratories is the wide variety of electronic data deliverable formats they produce. FDEP and SFWMD are in the process of implementing the ADaPT standard EDD format as a way to simplify and streamline the data review process. Once the standard EDD format has been adopted, EDDs from any contract lab can be imported into the ADaPT application for automated EPA level 3 data review. Reviewed and qualified EDDs can then be imported into the EDMS application, where the data will be combined with associated sample receipt and field information. EDMS can then compile information from multiple collectors into a single database and be made available to multiple users and agencies. Descriptions of the primary functions of both products follow below. The relationship between ADaPT and EDMS is illustrated in Figure 1.

Figure 1
ADaPT/EDMS Overview



Florida ADaPT

Florida DEP ADaPT is a computer application that processes two types of EDDs based on data elements needed for data quality assessment. These include the Laboratory Data Deliverable and the Laboratory Receipt Deliverable. ADaPT performs an error check for correctness and completeness on both the Laboratory Data Deliverable and Laboratory Receipt Deliverable. ADaPT also performs a data review on the Laboratory Data Deliverable that measures integrity of sample results against associated laboratory quality control, holding times, and method detection limits. The overall ADaPT process is summarized in Figure 2.

The Laboratory Data Deliverable contains sample result information including quality control batch links and accuracy and precision results for surrogates, laboratory control samples and matrix spike parameters. The Laboratory Receipt Deliverable contains sample information contained in the chain-of-custody. Both types of EDDs are constructed as a comma-delimited text files or Microsoft Excel .csv files and imported into the application for processing.

ADaPT uses a project specific library as the reference for EDD error checking and data review. The project library contains all the laboratory methods, analytes and their data review criteria such as reporting limits, blank contamination rules, holding times, accuracy and precision. A project library is created for each specific project. In this way, ADaPT has the flexibility to assess EDDs according to a particular project's requirements. ADaPT includes a Master Library containing a comprehensive list of nearly all possible methods and target analytes based on the State of Florida DEP standard values for method and analyte names. The Master library serves as a template for creating project libraries through an easy to use utility within the ADaPT program.

The ADaPT EDD error-check module examines both the Laboratory Data Deliverable and Laboratory Receipt Deliverable for correct standard values, missing information in required fields, format and duplicate records. Furthermore, the deliverable is also checked for target analyte completeness, consistent Lab Qualifiers according to the DEP laboratory qualification scheme, correct reporting limits, and the inclusion of results for various laboratory QC samples. After checking the EDD for errors, ADaPT creates an error log that can be viewed on screen or as a report. Each error is described in detail and, if applicable, the record number where the error occurs is identified. Laboratories have the ability to comment any or all errors if necessary.

ADaPT also includes a technical consistency utility that performs the following cross and reversal checks: nutrients (e.g., total nitrogen versus nitrate/nitrite and ammonia), total versus dissolved concentrations of the same sample, charge balance calculation, major ions vs. conductivity, TDS vs. conductivity, and analyzed TDS vs. calculated TDS. Technical consistency results can be viewed on screen or as reports. On screen view and reports can be filtered to show all results, anomalies only or cases where calculations could not be performed because of missing constituents.

Both EDD error-check and technical consistency modules are processes to be performed by the laboratory generating the data. ADaPT can and should be used by the laboratory to check the Laboratory Data Deliverable and Laboratory Receipt Deliverable and correct them as necessary before transfer to the client.

Automated Data Review evaluates the Laboratory Data Deliverable and appends data review qualifiers to sample results based on laboratory quality control information reported in the EDD and project specific data review criteria identified in the project library. Data review qualifiers and reason codes provide a coded explanation which shows specifically why result was qualified. ADaPT provides a variety of post data review qualification and outlier reports summarizing the results of the automated data review.

FDEP EDMS

EDMS is an Access application that processes three types of electronic data deliverables (EDDs) based on data elements needed for data quality assessment defined by the Florida DEP and. These include the ADaPT reviewed Laboratory Data Deliverable, the Field Data Deliverable and the Laboratory Receipt Deliverable. EDMS performs an error check for correctness, and completeness on both the Field Data Deliverable and Laboratory Receipt Deliverable. EDMS was developed as a tool for project managers to collect and process both field and laboratory environmental chemistry data. EDMS can generate reports that measure project completeness; compare results against historical information, contamination limits and preliminary remediation goals; and summarize analytical quality based on automated data review from ADaPT. EDMS also allows custom query generation for data retrieval and export to other third-party databases, such as FL STORET and SFWMD Water Quality Database (DB Hydro).

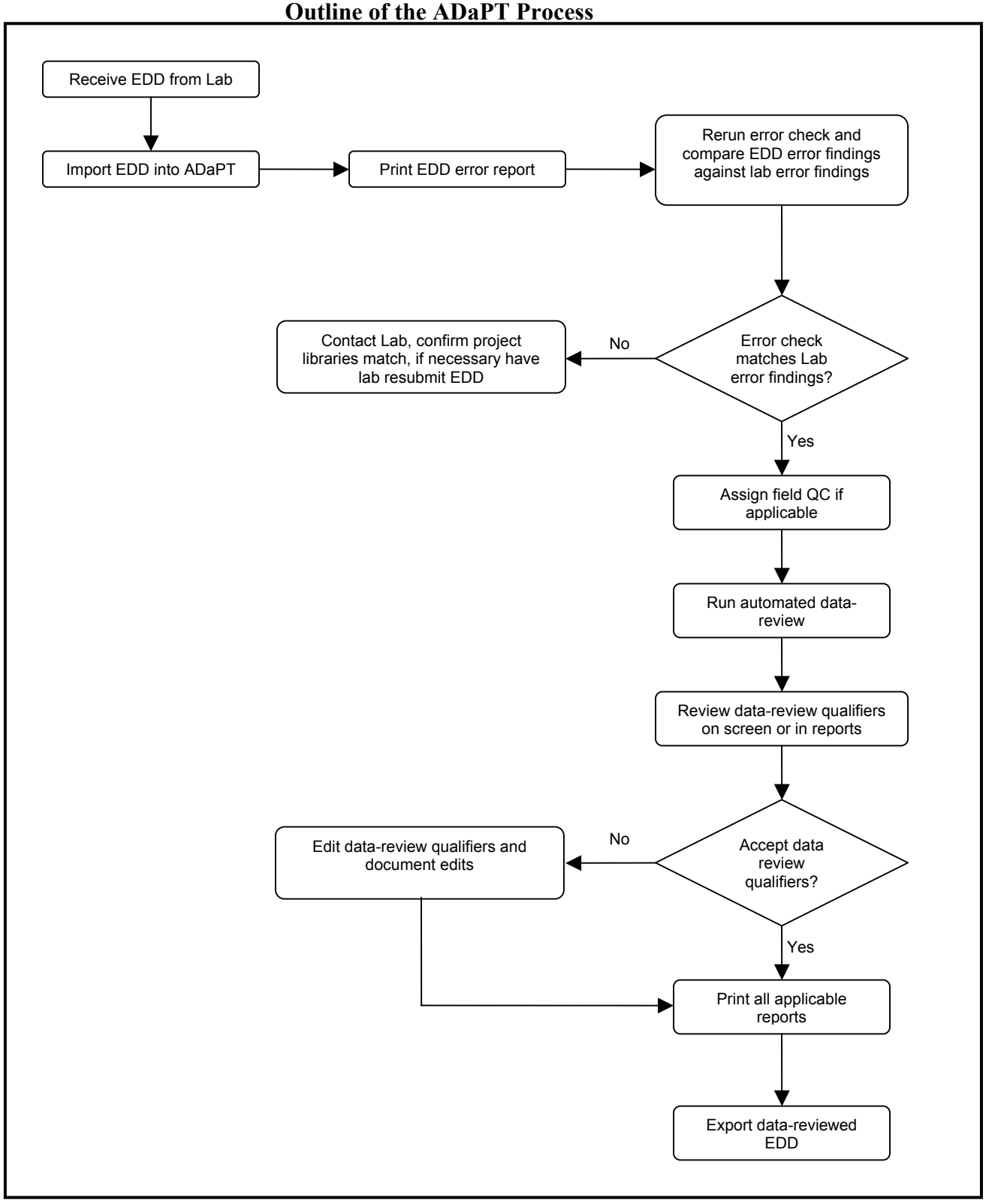
The Laboratory Receipt Deliverable contains sample chain-of-custody information. The Field Data Deliverable contains sampling, location, and field measurements information. The reviewed Laboratory Data Deliverable contains all the analytical information provided by the laboratory plus new information provided by the review/validation from ADaPT. The three EDDs are constructed as a comma-delimited text files or Microsoft Excel .csv files and imported into the application for processing.

The EDMS Field Data Management module examines both the Florida Field Data Deliverable and Laboratory Receipt Deliverable for correct standard values, missing information in required fields, format, logical values, and duplicate records. After checking the EDD for errors, EDMS creates an error log that can be viewed on screen or as a report. Each error is described in detail and if applicable the record number where the error occurs is identified.

Conclusion

Implementation of the ADaPT/EDMS process will significantly reduce the amount of time spent in manual review of data, improve the overall quality of the data coming from the contract labs and will standardize the format of the data to enable sharing of datasets among various agencies and contractors. This is a cost-effective method of improving both data quality and usability.

Figure 2
Outline of the ADaPT Process



The Impact of Electronic Record Regulations on the Pharmaceutical and Chemical Industries
Maximize profits or comply with regulations? Good data management software lets you do both.

Victoria Lander, NuGenesis Technologies

It's panic time. Your company is being sued over an environmental issue, and the company attorney wants all the site files from 1996 by the end of the week. The ring binders with the GC-MS printouts make a stack as tall as you are. Your geologist is on vacation this week, and you can't decipher his handwritten field notes. Not only that, but the lab tech mistakenly discarded the chain-of-custody sheets when the lab analyses were completed. You compiled all this data into a report for the district manager in 1998, but it got lost in the 1999 computer system upgrade. You go through a similar paper chase every time you have an environmental audit, your division manager is on your case to improve your productivity statistics, and you only have 24 hours in a day.

Chemists working in industries regulated by the Food and Drug Administration (FDA) and Environmental Protection Agency (EPA) are only too familiar with this hypothetical scenario. The pressure to reduce costs and improve profit margins competes with the need to comply with significant new government regulatory policies, creating an atmosphere of anxiety that can inhibit decision-making and the implementation of solutions.

The last decade has seen astounding growth in the generation of production, quality control, and laboratory data. Converting this data into meaningful information becomes difficult, however, when data is held in disparate, unconnected systems. Capturing data and transforming it into viable information not only ensures compliance with FDA and EPA regulations, but can enable companies to use this information with greater agility.

21 CFR Part 11

Industries that deal with food, drugs, cosmetics, nutraceuticals, or medical devices are subject to regulation under 21 CFR Part 11, the FDA's Rule on Electronic Records and Electronic Signatures (*www.fda.gov/ora/compliance_ref/part11*). This rule is based on GxP, the "good practices" applications. (GLP is good laboratory practices, GMP is good manufacturing practices, GCP is good clinical practices, etc.)

Part 11 has been "final and effective" since August 1997. Firms are expected to have their procedural and administrative controls in place by now. They must also have plans for upgrading their legacy systems (existing computer infrastructure) with the technical controls for Part 11, which address electronic record security, integrity, traceability, and the proper use of electronic signatures. These plans must be detailed and contain a

reasonable timeline and firms must show progress toward implementing them. Records covered by Part 11 deal with inventories, calibrations, preventive maintenance, validations, training, customer complaints, adverse events, and similar items. There is no determined date when all firms must be in total compliance with CFR Part 11, but the FDA has recently increased active enforcement of this rule. The basic intent of Part 11 is to ensure that scientists working in labs governed by predicate (GxP) rules keep certain records in electronic form, and that these records are trustworthy, reliable, and legally equivalent to paper records and handwritten signatures.

CROMERRR

Most of the EPA's data collection protocols were implemented years ago, before there was a clear appreciation of data quality principles and in the absence of standards that protected the trustworthiness of data. Most of the data currently collected by the EPA is acquired by regulated organizations and local agencies using vastly dissimilar collection and analysis methods. Lack of consistent standards not only increases the time and resources needed to review results, it can also lead to erroneous conclusions when incompatible data is compared incorrectly.

For example, in the mid-1980s, one environmental testing lab in New Jersey analyzed soil and water samples for priority pollutants using GC-MS. The instruments were not networked and there was no protocol for handling and long-term storage of the data. The paper printout of the analytical report was the protected data entity. It never occurred to management that electronic records would need to be protected for accurate and ready retrieval in a compliant environment. Today, anyone needing to re-examine and reprocess those unprotected, unsearchable tape backups would face quite a challenge.
Developed with 21 CFR Part 11 in mind, the EPA's proposed Cross-Media Electronic Reporting and Recordkeeping Rule (CROMERRR) will provide the legal framework for electronic reporting and record keeping under the EPA's environmental regulations (*www.epa.gov/cdx/cromerrr/propose/index.html*). The rule was proposed on August 31, 2001, and the public commentary period closed on February 27, 2002. Leave this out since the rule is currently languishing.

CROMERRR will apply to most, if not all, reporting and record keeping currently required of EPA-regulated organizations (currently regulated under GLP). These records include master schedules, protocols, standard operating procedures, and quality assurance inspection reports; the electronic documents will have the same legal and evidentiary force as their paper counterparts.
CROMERRR technical controls specify

- the ability to generate copies of records in human-readable and electronic forms,
- logical and physical protection against record compromises,
- secure computer-generated date/time stamps and audit trails,
- a means of retrieving records readily in the normal course of business, and
- a means of searching and retrieving archived records that preserve the context, metadata, and audit trails.

Comparing CROMERRR and 21 CFR Part 11

Although there are commonalities between the intent and technical and procedural controls for 21 CFR Part 11 and CROMERRR, they are two separate regulations, conceived and maintained by two different governing bodies. However, it is likely that some firms will be regulated by both CROMERRR and Part 11, since they currently comply with EPA and FDA regulations.

Both regulations establish the requirements for trustworthy electronically maintained records, substantiate electronic reporting requirements, and set up the functional capabilities of electronic record retention and document receiving systems. Unlike Part 11, CROMERRR establishes a Central Data Exchange (CDX) system for receiving e-records. EPA-regulated entities that use electronic systems to create, modify, maintain, or transmit electronic records must use procedures and controls designed to meet the minimum criteria to ensure that e-records are admissible in court to the same extent as previously kept paper records.

Neither CROMERRR nor Part 11 dictate the specific software and hardware needed to meet electronic record keeping requirements. Thus, industries can take full advantage of emerging technologies as long as e-record trustworthiness and reliability are preserved. The systems' users are responsible for understanding the requirements and adopting appropriate compliance controls into their overall business practices. **None whatsoever – both rules strive to maintain technological neutrality.**

Scientific Data Management Solutions

Regulated industries have commonly experienced difficulty in retrieving their data in a timely manner. In future response to CROMERRR, many EPA-regulated laboratories may compile comprehensive reports for electronic submissions to the EPA. Like their pharmaceutical counterparts, these laboratories are challenged by their need to coordinate disparate data from a wide variety of sources. This data is ideally assembled from multiple sites, collected, archived, and used at one convenient location. Scientists might also be called on to locate and readdress the data even after they have submitted their final reports.

Developing a regulatory compliance strategy is not an easy task. Many software suppliers have enhanced their product's features and devised internal technologies to address the technical controls required by these regulations. Such fixes tend to be proprietary and thus become only partial solutions. Any regulation intended to protect the reliability of records generated by computer-controlled systems not only relies on the regulatory technical controls, but also on the implementation of procedural and administrative controls. These are the policies and procedures that govern conformance to the technical requirements of a regulation—for example, a written procedure prohibiting users from sharing their passwords for system access.

In practical terms, a thorough approach to data and knowledge management requires a flexible, Web-based platform that integrates with multiple analytical and software systems to provide an enterprise-wide record keeping solution. Such a platform must also incorporate the technical controls for e-record maintenance and e-signature regulation. This information must be stored in a database that is compliant, easily retrievable, and reusable.

Pulling It All Together

One large pharmaceutical company learned firsthand how such a data management system could help them make the most of their data. This organization was interested primarily in increasing its ability to meet the Part 11 requirements for raw and report laboratory data. In the existing system, report data was stored on paper, and raw data files were deleted from analytical instrument computers when free disk space was required. It was difficult to find data when it was needed, and the company was not protecting its raw data for retrieval throughout the file retention period required by Part 11. The company implemented the NuGenesis Scientific Data Management System (NuGenesis Technologies Corp., Westborough, MA), a program that unifies data from various sources into a common Web-based electronic format.

By installing, validating, and using this software along with the appropriate Part 11 procedural and administrative controls, researchers were able to capture the raw analytical data from their HPLC, LC-MS, and other instruments, and store human-readable report data in a secure relational database. Metadata was extracted automatically and cataloged, enabling successful data searches. More importantly, access to the data was controlled by system logins, automatic computer-generated time-stamped audit trails, and other security features. At this point, the researchers in this laboratory felt that they were in 90–95% compliance with Part 11, a significant improvement over the previous system.

CROMERRR has not yet been finalized, but because the record keeping requirements for CROMERRR are based on those already in place for 21 CFR Part 11, the software requirements will be similar for both sets of regulations, avoiding the need for two separate systems.

The general feeling in the industry concerning these rules on trustworthy record-keeping is that although implementation of their requirements is costly and time-consuming, when all of the dust settles, these regulations will pave the way to more reliable data management and utility. Trustworthy records in our current environment of hackers and manipulators wll undoubtedly add an element of confidence to the regulated arenas.

Suggested Reading

Information on 21 CFR Part 11 and CROMERRR;
http://pw1.netcom.com/~jlboet/esiglinks.html.

NuGenesis Technologies' 21 CFR Part 11 page; www.21cfrpart11.com.

Society of Quality Assurance CROMERRR updates;
www.sqa.org/Committees/CROMERRR/CROMERRR.htm.

U.S. EPA Central Data Exchange;
http://alpha.lmi.org/epa/electronic_reporting/cromerr.html.

Victoria Lander *is the market manager for compliance for NuGenesis Technologies Corp. Send your comments or questions about this article to success@NUGENESIS.com*

Measures of Quality System Implementation

Louis Blume[1], U.S. Environmental Protection Agency, Great Lakes National Program Office, 77 West Jackson Boulevard, Chicago, Illinois 60604, Judy Schofield and Molly Middlebrook, DynCorp (A CSC Company), 6101 Stevenson Avenue, Alexandria, VA 22304

Quality programs are not implemented with the stroke of a pen upon the approval of a Quality Management Plan. They take time and are built through a process of continuous improvement that identifies mistakes, implements corrections, and promotes continual, honest self-evaluation. Functional quality programs do not just happen - they evolve. Metrics can be used to quantify various degrees of quality implementation. This paper presents a proposed metric and reviews the key steps of quality implementation. The paper also emphasizes why quality system reviews are a true benefit and not an inquisition.

The ability to recognize the varying degrees of implementation of quality systems is critical to the successful long-term adoption of value-added quality programs. Too often, overly optimistic expectations, or characterizations in the infancy stages of development, tend to derail programs that have the potential to make a significant difference over the long term. The quality system of the US EPA's Great Lakes National Program (GLNPO), over its tumultuous 12-year history, is used as a test case, while discussing the numerous and varied external organizations to which GLNPO provides oversight. We apply the proposed metric for quantifying the degree of quality implementation to these various programs and discuss its utility. This paper should provide some level of hope to eager, energetic, aspiring Quality Managers who are frustrated by the short-term lack of enthusiasm, attention, or progress of their programs.

PERFORMANCE-BASED MANAGEMENT OF THE ENVIRONMENTAL RESTORATION PROGRAM

Dr. Javier Santillan (AFCEE)
Mr. Jim Gonzalez (AFCEE)
Dr. Marc Gill (BAH)

The US Air Force has noted that optimization is not the universal answer to speedy remediation. This is due to the numerous uncertainties inherent to site remediation. Identification of all restoration uncertainties is required to manage and minimize those uncertainties. USAF guidance on Performance-Based Management OF THE Environmental Restoration PROGRAM was prepared to aid project managers in handling restoration program uncertainties. These uncertainties include:

- *Accuracy of the Conceptual Site Model*
- *Adequacy and attainability of the remedial action objectives (RAO)*
- *Effectiveness and efficiency of the selected restoration technology*
- *Adequacy of the metrics selected to monitor performance and effectiveness*
- *Risks and pollution created by the restoration activities*

Restoration programs based on optimized, performance-based, remedial and monitoring processes have a high probability of success. However, the measure of "success" must be defined in terms of stakeholder approved performance-based metrics. A cleanup plan using decision logic should state how cleanup goals would be measured to demonstrate their attainment, and a completion plan can be used for this purpose. Completion plans should be updated whenever new information is available.

The Office Of Pesticide Programs' Quality System

Betsy Grim
Director of Quality Assurance
Office of Pesticide Programs
United States Environmental Protection Agency
Washington, D.C.

The EPA's Office of Pesticide Programs' (OPP) mission is to protect human health and the environment from unreasonable adverse effects from the use of pesticides. OPP is the largest program within EPA and is unique because it is a licensing program. In order to ensure that all environmental measurements including secondary data supported by the OPP are of known and acceptable quality and can readily be used to support regulatory decision-making, OPP has a Quality Management Program that creates an environment that continually evaluates and improves work processes and products. The OPP Quality Assurance (QA) Program is documented in a Quality Management Plan (QMP). The QA Program is reviewed internally and externally to assure that standards documented in the Quality Management Plan are met.

What Are Data of Known and Documented Quality?

Marlene Moore
Advanced Systems, Inc.
Newark, DE

The accreditation of laboratories ensures the generation of data of known and documented quality. This has led to confusion and misunderstanding within the environmental user community. What does it mean when a laboratory is accredited and therefore is expected to generate data of known and documented quality. This session will present what is meant by data of known and documented quality as it relates to the generation of data under today's accreditation standard (2001 NELAC) and the upcoming approved NELAC (National Environmental Laboratory Accreditation Conference) standards adopted in 2002 and 2003.

Introduction

The laboratories currently accredited under the 2001 NELAC standard must demonstrate capability of the method and analyst and have documentation of this demonstration on file during the assessment. Although this does not change in the 2002 and 2003 standard for documentation of analyst demonstration of capability it will significantly change in the 2002 and 2003 standard for the documentation of method uncertainty (2002) and for method performance for chemistry (2003). The requirements for documentation of the demonstration of method performance is defined in the 2002 standard as requiring method validation and determination of the estimation of uncertainty. The 2003 standard provides additional requirements for the demonstration of method performance for Chemistry (D.1) Radiochemistry (D.4), Air Testing (D.5)\ and Asbestos (D.6). The 2003 standard add the definition for the documentation of method performance for the limit of detection (LOD), the limit of quantitation (LOQ), precision, bias and selectivity.

How do these changes affect the use of the data for making environmental decisions? The most significant difference in the 2002 is that the laboratory must ensure that the data is appropriate for the intended use by the client and understanding the contributors to the uncertainty of the measurement. This process is dependent on the client needs, laboratory service definition or the regulatory definition for the data. An accredited laboratory (after adoption of the 2002 standard) must determine the use of the data and work with the client to be sure the data is appropriate for the intended application. The reporting of the estimated uncertainty for the measurement provides the data user with information of the variability of the result reported. The reporting of method deviations and qualification of performance data provides additional information on the data quality.

It is often misunderstood that the data from these accredited laboratories is appropriate for any intended use by an EPA program or other decision maker. Data generated for one purpose may not be appropriate for use in making a different decision at the same site. Data of known and documented quality means the laboratory has defined under what conditions the data were generated and has available all qualifications on the use of the data. As NELAC expands to accreditation of field activities, it is realized that these same qualifications are also required for field activities such as sample collection and field-testing.

Discussion

The following provides an overview discussion of the more significant differences in the method performance documentation requirements for the three NELAC standards. This does not represent all the changes and the specific standards should be referenced for details on the requirements. Each of the following discussions identifies the section of the NELAC standard to find the requirements.

Method selection, validation and measurement uncertainty

The adoption of the ISO/IEC 17025 requirements into the 2002 NELAC standard added the requirements for method selection, validation and estimating the uncertainty of the measurement. These elements have resulted in considerable discussion, which resulted in the changes reflected in the 2003 standard.

The ISO/IEC 17025 enhances the requirements from the 2001 standard, by focusing on the intended use of the data and making sure the method selected is appropriate for the intended use. In addition it requires that the method when it is used for this purpose is appropriately validated. The degree of rigor necessary for the laboratory to validate the method is based on client needs and the intended use of the data. For example, mandated methods provide method demonstration or performance criteria. These methods are selected when the test data are used for reporting under a specific regulatory requirement. The method validation is limited to documentation of meeting the performance criteria specified by that mandated method. Batch quality control is usually specified by the regulated method and if not the additional QC required by NELAC is added by the laboratory to ensure documented data has the same QC to allow comparability of the data. The 2003 standard provide more specific direction for defining method validation criteria by requiring documentation of the limit of detection (LOD), limit of quantitation (LOQ) precision and bias demonstrations.

The estimation of uncertainty is defined in the ISO requirements and was adopted by NELAC in 2002. This term is defined by the international community to allow representation of all the factors that contribute to the variability of the measurement to be expressed in terms of an interval about the result reported (e.g.; a plus/minus value). The mathematical model for representing this is defined in international documents, but is not widely accepted within the environmental community. The calibration community, industrial community and others such as food are adopting and implementing this international definition.

The estimation of uncertainty when applied to environmental measurement (e.g.; Visual Sample Plan (VSP)) provides the decision maker with the necessary information relative to range of values where the estimated true value lies. The estimation of uncertainty in the laboratory can only be related to the contributors from the laboratory method such as the bias, precision, and other factors that determine the correctness and reliability of the method as performed by the laboratory on the specific sample container received. The measurement uncertainty to represent the estimate for the site or project can only be developed when the client, sampling organization and laboratory work together to optimize sampling and test method performance for the specific project or sample matrix being tested. (2002 Section 5.5.1)

Calibration

The NELAC standard has always required the bracketing of the sample results with calibration standards. NELAC laboratories must qualify data as estimated when reported sample values fall

outside this range. The 2001, 2002 and 2003 do not change this overall intent, but each year the standard has been updated to clarify and provide more specific requirements to define this bracketing of the sample results requirement.

The 2002 NELAC standard added clarification to the requirement for calibration of instruments when the manufacturer specifies a blank and single standard calibration (such as ICP, ICP/MS). In addition to the calibration using the single point, the laboratory must demonstrate that the instrument can measure concentrations at the lowest quantitation level. The zero, calibration standard and lowest quantitation level standard must be analyzed each analytical batch (e.g. samples analyzed as a group). (2002 Section 5.5.5.2.2.1.f)

The 2003 NELAC standard (Section 5.5.5.10) removed the requirement to vary the continuing calibration standard, but more clearly specifies the requirements that a calibration standard must be at the lowest concentration and highest concentration for which quantitative data are to be reported. Any data outside these values are to be qualified or explained in the case narrative. Where instruments use only one calibration standard, the linear range must be established and a limit of quantitation (LOQ) standard must be analyzed with each analytical batch. This LOQ must demonstrate to be within acceptable criteria. It is noted that there are currently different definitions for "acceptable criteria" for this part of the standard. The standard does not clearly indicate if these acceptable criteria must be the same as other points within the linear range of if this can be wider.

For critical decision making the data user should be sure that the action limit and the limit of detection or quantitation is within defined acceptable limits for the data use. Otherwise the laboratory may set default limits to demonstrate performance, but this performance may not be adequate for the final decision making such as a 50% recovery at the LOQ. The requirements for calibration of equipment and the specific information on when data are to be qualified must be defined in the project specific or regulatory specific criteria to ensure adequate quality for the decision. It is not up to the laboratory or NELAC to specify the acceptance criteria (e.g; 20%, 30% 50% recovery, percent difference or RPD).

Blanks

The requirements for the measurement of the blanks are consistent in 2001 and 2002 standards. The 2003 standard (Section D.1.1.1) adds the requirement that the laboratory must take and document corrective action when the blank is defined as contaminated in order to minimize or eliminate the contamination. Blanks are analyzed once per preparation batch or one per twenty samples.

Data users must define specific criteria for qualifying blank data when it is below the quantitation limit but found in the blank. If the action limit or decision level is above the quantitation limit the need for reporting blank values below the quantitation data is not necessary and the qualification of the sample data in these cases should not be required.

Detection Limit

In the 2001 and 2002 standards, the laboratory must define the procedure for determining the detection limit for the method. In most cases 40CFR Part 136 dictates the procedure since most states mandate the use of this method for determining the method detection limit (MDL) for all programs (drinking water, wastewater, solid waste). The frequency for making this determination is specified by the mandated method or is performed as defined by the laboratory.

Many of the mandated methods do not define an annual frequency, but state laws or regulations mandate the frequency of annual determination. The mandated method or regulatory program often takes precedence as required in all three of the NELAC standards.

The term Limit of Detection (LOD) first appears in the 2003 NELAC standard (Appendix C.3.1). The American Chemical Society (ACS) has used this term in the past. The LOD is "an estimate of the minimum amount of a substance that an analytical process can reliably detect. An LOD is analyte-and- matrix-specific and may be laboratory dependent." (NELAC 2003 Glossary). The LOD must include all sample processing steps for the method and all analytes reported for that method. The LOD value is confirmed by qualitative identification of the analytes in the quality system matrix. The concentration must be no more than 2-3 times the LOD for a single analyte method or 1-4 times the LOD for multiple analyte tests. The verification of the LOD must be on each instrument used by the laboratory to report this value. If the LOD study is not performed then no results are to be reported below the limit of quantitation.

Laboratories are required to follow the mandated methods, regulatory requirements and NELAC. Therefore after the effective date of the 2003 NELAC standard in July 2005, laboratories may be performing MDL studies and LOD verifications for a number of methods. The laboratory must have a procedure to define how the LOD relates to the LOQ. But the 2003 NELAC standard does not require this relationship to be extended to the MDL.

Data users must clearly indicate the need for detection limit studies and work with the laboratory to ensure that the detection limit data is relevant to the intended use of the data.

Quantitation Limit

The 2001 and 2002 NELAC standards indicate that the quantitation limit is the lowest standard on the calibration curve. The 2003 standard more clearly defined the determination and demonstration necessary to define the quantitation limit used for reporting sample data.

The term Limit of Quantitation (LOQ) first appears in the 2003 NELAC standard. ACS used this term in the past. The LOQ is "the minimum levels, concentrations, or quantities of a target variability (e.g. target analytes) that can be reported with a specified degree of confidence." (NELAC 2003 Glossary) The 2003 standard requires the verification according to the laboratory procedure of the LOQ at 1-2 times the claimed LOQ. The LOQ must be within the test method acceptance criteria or client data quality objectives for accuracy. This single analysis is not required if the bias and precision measurement system is evaluated at the LOQ.

Data users must clearly indicate the level necessary for quantitation to ensure the quantitation limit is appropriate and relevant to the intended use of the data.

Laboratory Control Sample (LCS)

The 2001 and 2002 NELAC standard (Sectrion D.1.1.2.1) modified the requirements for the laboratory control sample (LCS). The requirements are to perform the LCS one per preparation batch. All targeted compounds are to be in included in the LCS over a two year period whenever the mandated method does not specify the list of spike compounds. For example, solid waste methods performed per SW-846 method 8260B include five spike compounds for the GC/MS volatiles methods. Therefore the need to spike all compounds in the LCS or MS/MSD is not required per NELAC. In methods where the spike compounds are not specified then all compounds must be spiked according to a defined number of target compounds. When less than 10 target compounds are analyzed, all 10 target compounds must be in the LCS. When 11 to 20

targets are analyzed the laboratory must spike at least 10 or 80 % per batch and greater than 20 targets requires at least 16 spike compounds.

The evaluation criteria in the 2001 and 2002 standard required all these spike compounds to be acceptable. In the 2003 standard this is changed to allow a specified number of compounds to exceed the control limit (3 times the standard deviation) for long lists of analytes (greater than 11) but must be within the marginal exceedence limit (ME). The ME is 4 times the standard deviation around the mean. For example if 31 to 50 analytes are in the LCS then two analytes are allowed to be within the 3 and 4 times the standard deviation of the mean. If any LCS analyte exceeds the ME then the LCS fails and the data must be qualified. (Section D.1.1.2.1.e)

Data users must understand the NELAC standard quality control criteria for the LCS to determine if this information is adequate for the decision. In some cases it may be in the data users best interest to specify that the laboratory for a given project adopt specific QC requirements from the NELAC standard especially when the NELAC requirements are more stringent than the mandated method. For example spiking all target compounds in the LCS or MS rather than a limited list of compounds as specified in the reference method.

Precision, Bias and Selectivity

The 2001 and 2002 standards require the use duplicates to evaluate precision and the LCS to evaluate the bias of the laboratory. The matrix spike is used to evaluate the method performance on a client sample, but is not used to evaluate laboratory performance of the method.

The requirements for demonstration of precision, bias and selectivity are found in the 2003 standard. Demonstration of the precision and bias for standard methods must be performed as required in Appendix C.1 and is consistent with the current NELAC standard requirements. The demonstration of precision and bias for non standard methods is defined in Appendix C.3.3.b. The laboratory must have a procedure for evaluating the precision and bias for these nonstandard methods. The laboratory performance results are to be compared to client requirements, reference method requirements or criteria established by the laboratory. This demonstration must be performed across the calibration range and on each quality system matrix. The samples used for the performance demonstration must be processed through the entire analytical process including all preparation steps for each analyte of interest. The approach to be used for this evaluation is based on the laboratories systematic approach. Two examples are presented in the NELAC standard but are not required. The precision and bias information from the laboratory studies provides the data user with the laboratory specific performance of these nonstandard methods.

The selectivity requirements in the 2001 and 2002 standard are limited to three items, retention time windows, mass spectral tuning and confirmation of the identification of positive results for samples where the compound has not been verified as present, such as organic analysis by gas chromatography. The 2003 standard (Section C.3.4) provides more general guidance that requires the laboratory to establish checks for more than the three items found in 2003 Section D.1.5. These additional items include ICP inter-element interference checks, spectrochemical absorption or fluorescence profiles, co-precipitation evaluations and electrode response factors.

Data users must specify the bias, precision and selectivity requirements to ensure the limits and selectivity is appropriate and relevant to the intended use of the data.

Other

Other quality control (QC) requirements are found in the mandated methods or may be specified by the client. The NELAC standard specifies that the laboratory must assure the quality of the results by having essential QC elements (2002 Section 5.5.9) and sample specific QC measures (2002 Section D.1.1.3). These include the use of surrogates, matrix spikes, duplicates and those QC requirements specified in the mandated method, the laboratory's program and as required by the client such as project specific requirements defined in a QAPP. The sample specific controls are not used to assess laboratory performance but are used to assess the performance of the method on the specific sample matrix. The NELAC laboratory must qualify this data and report when the matrix does not demonstrate acceptable performance using the selected method.

Conclusion

Data of known and documented quality means the laboratory has defined under what conditions the data were generated and provides data qualification on the use of the data. Accreditation provides the data user with the assurance that the laboratory has the equipment, personnel, and recordkeeping process and management systems to generate the data within a continuous improvement operation. Accreditation cannot guarantee that the data is appropriate for the intended use unless the client and laboratory work together to ensure the method selected is suitable for the samples being analyzed for that client and the application of the data by the decision maker.

References

"EPA Requirements for Quality Assurance Project Plans", EPA QA/R-5, EPA/240/B-01/003, March, 2001, Office of Environmental Information, Quality Staff, U.S. Environmental Protection Agency, Washington, DC.

"EPA Guidance for Quality Assurance Project Plans", EPA QA/G-5, EPA/600/R-02/009, December 2002, Office of Research & Development (ORD), U.S. Environmental Protection Agency, Washington, DC.

"Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs." American National Standard, ANSI/ASQC E4-1994

National Environmental Laboratory Accreditation Conference (NELAC) standard adopted July, 2001, 2002 and 2003

Qualifying Quantification of Measurement Variability

Marcus E. Kantz, U.S. EPA, Region 2

Quantifying Measurement Variability is like handing vegetable soup to your friend - in your hand. Your friend may well get all of the vegetables, but none of the soup. In other words: What you can quantify, you can quantify well, but what you can't quantify, you probably can't quantify at all, and a handful of vegetables is not soup. There is currently a major effort to quantify the variability in environmental measurements. This is a great idea, ultimately because it would improve environmental decision-making. However, it is almost impossible, and the mere effort gives a false sense of security. Each step in the measurement process comes with its own inherent variability: the environment itself, sampling, analysis, and data handling. We can develop a pretty good handle on analytical and data handling variabilities, but those associated with the environment and sampling are much more difficult. The big problem is that the variabilities accumulate along the process. Since the biggest variabilities, and those that we can quantify the least, show up first, they dominate the process. The biggest problem is that the world is not homogeneous, and it changes all the time. We can try our best to sample the environment at critical places and critical times, but we never know for sure how successful we are. The most important conclusion then, is: Don't mislead yourself into thinking that you understand the variability associated with environmental measurements. Pretending you understand will just give you a handful of wet vegetables. Acknowledging the truth is much closer to soup.

Quantifying Measurement Variability is like handing vegetable soup to your friend - in your hand. Your friend may well get all of the vegetables, but none of the soup.

In other words: What you can quantify, you can quantify well, but what you can't quantify, you probably can't quantify at all, and a handful of vegetables is not soup.

Before we can look at variability in measurement, we need to understand the story of the measurement process. The goal of environmental measurement is to quantify our understanding of the environmental condition at the time and place and situation of interest in order to provide factual evidence for making some decision. So, **useful** understanding is soup, numbers are just the vegetables.

Our problems in understanding our measurements start with the environment itself. The environmental measurement process is not perfect. The result is always uncertain to some degree because of variability in the environment itself, variability in the measurement process, and, of course, error in the measurement process. This error is, in turn, due to imperfect understanding of the measurement process and imperfect execution of the measurement process. This is then compounded by our inability to understand and quantify the above.

There is currently a big push to quantify the variability in environmental measurements to better support decision-making. This is a great idea, since it can reduce uncertainty and improve our chances of making the right decision. However, it is simultaneously a dangerous idea, since it can give a false sense of security, as well as a false answer. The goal of this presentation, then, is to highlight the hidden traps in attempting to quantify environmental measurement variability.

The 5 sources of variability in the environmental measurement process are:

- The Environment
- Sampling
- Laboratory analysis
- Data handling
- Our understanding of the first 4

Our job is to investigate the problems in trying to calculate the variability encountered at each step and to combine them all, while accounting for the propagation (and, thus expansion) of variability as the measurement process proceeds. We'll use a simple graphical tool, charting the cumulative variability as we move along the "typical" environmental measurement process. We'll base all of our estimates on the "true value" of our arbitrary measurement exercise, and see how far we stray from it as the variability progresses. As you can see in Figure 1, this true value of whatever we're measuring is represented by a horizontal line as the measurement process marches left to right from start to finish. Meanwhile, any variability above or below this true value would be shown above or below it as the process moves on. Figure 2 shows the terminology we'll use to describe the four main steps in the process: E, S, A, and D.

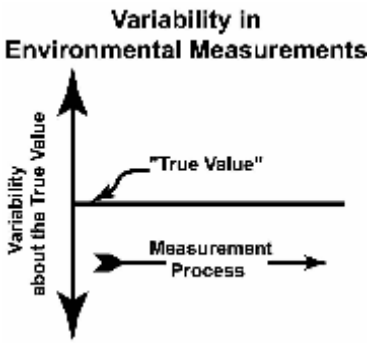


Figure 1

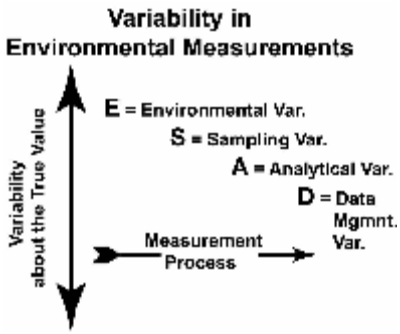


Figure 2

Using this tool, we'll look at each step in the measurement process, indicating the variability associated with it, starting with the Environment.

Environmental variability is not predictable, and is only partly testable. As it affects measurement, it includes spatial and temporal factors, is usually affected by both natural and anthropogenic factors, and is very complex. The world is way heterogenous, and it changes. We can't change that. There are ways to test, or at least estimate the natural variability through thoughtful sample design (see QA/G-5S). But we can't "minimize it" and we shouldn't try. It is what it is. The green triangle in Figure 3 shows how we would like to picture the magnitude of environmental variability, starting at the true value and increasing as the environment changes from time to time or place to place. The trouble, though, and the reason there is a red line through it, is that **there is no single true value in nature**, There is only a range of true values that exists before we decide to measure, and continues after we are done, as shown in Figure 4.

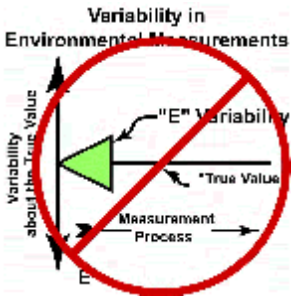


Figure 3

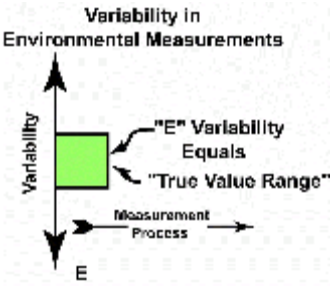


Figure 4

Now we can move on to the variability associated with sampling. Sampling variability is not predictable, and is only partly testable. It also includes both spatial and temporal factors, and seems to be "man-made." But, sampling variability is heavily dependent on how well you understand the environmental variability. This connection is called Representativeness.

It can be very, very hard to get a sample of what you want to measure (Remember, at this point the environmental variability has already been accounted for by the green bar. It's a given.). There are ways to minimize variability due to sampling (such as use of a spatially and/or temporally dense sampling network, and lots of replicates). There are also ways to improve sample representativeness, and thus reduce variability due to sampling (such as through the use of statistical sampling design, as described on the EPA QA web site and in G-5S). But it is still hard to collect representative samples, and even harder to know how well you have done it. Statistical sampling design can add testability of sampling variability, but it may still be hard to differentiate between environmental and sampling variability.

As you see below, we will treat the accumulation or propagation of variability as additive. In some cases, it **is** additive, as when two consecutive variabilities are as ± x and ± y ppm then the combined variability is ± (x + y) ppm. Additive variability is, in fact, the minimum accumulation rate. There may be cases in which the variabilities are multiplicative, or even higher. So remember, the graphical descriptions are a **best case**, not a worst case. They're just easier to draw. For the same reason, we will show the variabilities as symmetrical, plus and minus, but they need not be symmetrical.

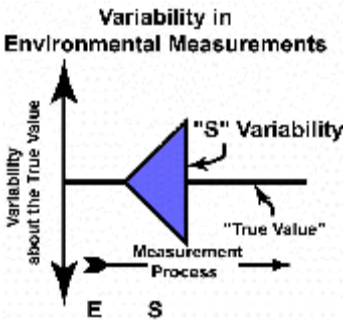


Figure 5

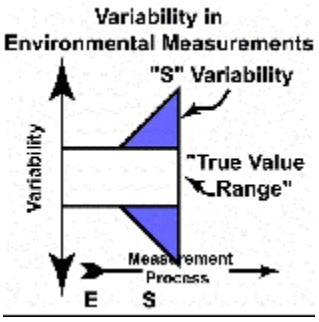


Figure 6

Analytical variability is somewhat predictable, or at least testable. It includes both the variability inherent in the chosen method along with any variability associated with the lab's implementation of it. 'Typical' variability of the method may be well documented. The lab's individual variability is testable during the project by including appropriate QC checks in the process (blanks, duplicates, spikes, PTs, etc.).

Data handling variability is not predictable (computers are perfect, but their programmers are not), but it is testable. It includes the effects of transcription, telemetering, conversions, etc., and is almost entirely due to human error. (Note that this does not include variability associated with modeling. That is a shoe of a different color.).

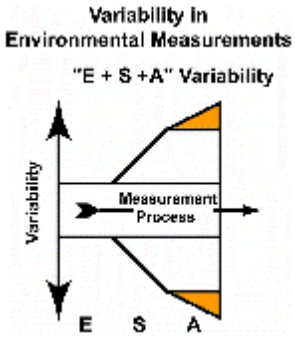


Figure 7

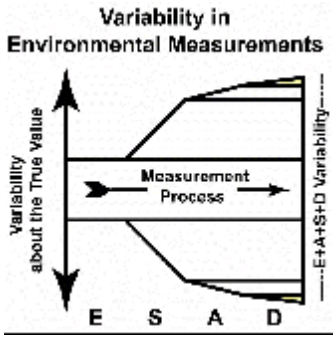


Figure 8

Our understanding of environmental, sampling, laboratory, and evaluative processes are all incomplete, inaccurate, and insular. The variability of an environmental measurement process is the result of the interrelationships of all of these. They occur sequentially, and they accumulate sequentially. Any lack of understanding of one of the earlier variabilities results in a much greater lack of understanding of the cumulative variability as the estimation process continues. And the ones we understand the least are the first ones. Too bad for us.

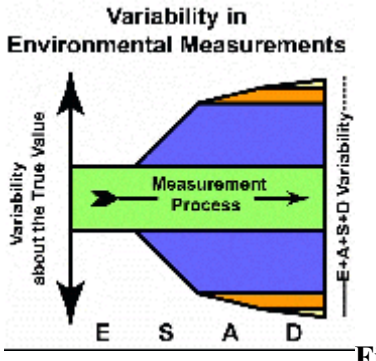


Figure 9

In conclusion, we must start with the acknowledgment that variability is a part of life. We can reduce it, but not eliminate it. We can sometimes measure or estimate it. Most importantly, in environmental measurements: don't mislead yourself. You can't completely quantify the variability of the environmental measurement process. Pretending you can, is just vegetables. Acknowledging the truth can lead to more enlightened decision-making, and can be soup.

Cost Effective "Collaborative Sampling" in Visual Sample Plan (VSP) Software to Estimate Means and Test Hypotheses

Richard O. Gilbert, Brent A. Pulsipher and John E. Wilson
Pacific Northwest National Laboratory
Richland, Washington

Selecting a cost-efficient sampling design for determining the right type, number and location of environmental samples is a critically important component of any environmental study. This paper presents an innovative design called Collaborative Sampling (CS) that can be more cost-effective in some situations than simple random sampling. The CS design uses two measurement methods: a field-based relatively inexpensive measurement method, and the standard laboratory "expensive" method. The idea behind CS is to replace the need for obtaining so many expensive measurements with collecting a larger number of the less expensive measurements. The CS design is currently being added to the suite of designs in the Visual Sample Plan (VSP) software, which can be downloaded free at http://dqo.pnl.gov/vsp. This paper discusses the CS methodology, assumptions and VSP implementation of the CS design.

1.0 Introduction

The importance of selecting a sampling design for obtaining representative environmental data for decision making cannot be disputed. The application, benefits, and limitations of several basic and innovative sampling designs are discussed in EPA (2002). The Collaborative Sampling (CS) design, which is not discussed in EPA (2002), can be more cost effective in some situations than simple random sampling for estimating the mean and testing hypotheses about the mean. Although CS may be new to many environmental professionals, discussions of the CS design can be found in, e.g., Gilbert (1987) under the title of "Double Sampling."

The CS design uses two measurement methods: the "standard analysis" (sometimes called the laboratory analysis or "the expensive method") and a less expensive and possibly less accurate measurement method (sometimes called the field-based analysis or "the inexpensive method"). The idea behind CS is to replace the need for obtaining so many expensive measurements with collecting a larger number of the less expensive measurements. The inexpensive method is used at n' locations and the expensive method is used at n of those n' locations, where n' is typically much larger than n.

The CS design is currently being implemented in the Visual Sample Plan (VSP) software tool for use when the sampling objective is to estimate a mean, compute an upper confidence limit on the mean, or test whether the mean exceeds an upper threshold value. VSP is map-based, user-friendly visual tool that helps the user determine the number and location of samples needed to ensure confident decisions. It is focused primarily on sampling design but some modules, including the CS module, incorporate statistical analysis routines for analyzing the data once it has been gathered.

2.0 Estimating the Mean

Suppose the objective of sampling is to estimate the mean of a contaminant in surface soil over a defined geographical region. One design that might be considered is simple random sampling (or perhaps systematic grid sampling) to select sampling locations, and then use the standard ("expensive") laboratory analysis method on the collected samples. Should the CS design be used instead? As discussed in Gilbert (1987, Chapter 9), the following conditions must hold for CS to be more cost effective than using the entire measurement budget to obtain expensive measurements on samples collected using a simple random sampling design:

- There is an underlying linear regression relationship between the two types of measurements
- There is a sufficiently high correlation, ρ, between the two types of measurements made at the same locations
- The ratio $R = C_{ex} / C_{inex}$ is sufficiently large, where C_{ex} is the cost of a single expensive measurement and C_{inex} is the cost of a single inexpensive measurement.

When the objective is to estimate the mean, CS will be more cost efficient than simple random sampling if the following inequality holds (Gilbert 1987, equation 9.5):

$$\rho^2 > \frac{4R}{(1+R)^2} \qquad (1)$$

In practice, the value of ρ will be uncertain and should be estimated using a "pilot" study in which the proposed inexpensive and expensive measurement methods are used in realistic field and laboratory conditions for, say 20 or more locations. Also, these pilot study data should be plotted in a regression scatter plot to assess the linearity assumption.

If CS is cost effective, then equations in Gilbert (1987, page 109) can be used to compute the number of samples, n' and n, needed. Gilbert provides equations for two cases:

- Minimize the variance of the estimated mean for a given fixed measurement budget
- Minimize the total measurement cost subject to the constraint that the variance of the estimated mean is no greater than the variance of the mean that would be obtained based on n expensive measurements obtained using a simple random sample design.

The above methodology (testing for cost efficiency and computing n' and n when the sampling objective is to estimate the mean) can be easily accomplished using the VSP software code. After booting up VSP, simply click on **Sampling Goals > Estimate the Mean > Data Not Required to be Normally Distributed > Collaborative Sampling > Simple Random Sampling** or **Systematic Grid Sampling** to access the dialog box for inputting the required Data Quality Objectives (DQOs).

3.0 Confidence Limits on the Mean

Suppose the sampling objective is to estimate the mean and also compute a one-sided upper or lower confidence limit or a two-sided confidence interval on the mean. In additional to providing an interval within which there is confidence the true mean lies, an upper confidence

limit on the mean is sometimes used to test if the mean exceeds a threshold value. A method for computing the required n' and n samples when the objective is to compute confidence limits has recently been developed by the authors and is currently being incorporated into VSP. The VSP user gains access to this method in VSP by clicking **Sampling Goals > Construct Confidence Interval on the Mean > Can Assume Data will be Normally Distributed > Collaborative Sampling > Simple Random Sampling** or **Systematic Grid Sampling**.

This CS module works very similarly to the CS VSP module discussed in Section 2.0 above. First, the VSP user inputs the following DQOs into the VSP dialog box (the desired width of the confidence interval, the desired confidence level, the expected total standard deviation of the data set of expensive measurements, the expected correlation ρ between the inexpensive and expensive measurements, and costs C_{ex} and C_{inex}). Then VSP determines if CS is cost effective using Equation 1.0 above.

If CS is cost effective, then VSP computes n' and n such that the total measurement cost, C, is minimized subject to the constraint that the width of the confidence interval (CI) will be no greater than a CI width that would be obtained using n_v samples obtained using simple random sampling and measured using only the expensive measurement method. This value of n_v is computed using an iterative procedure (Gilbert 1987, page 30). Then n' and n are computed using n_v and Equations 9.8, 9.9 and 9.10 in Gilbert (1987, page 109).

After the n' and n measurements have been obtained, the VSP user can enter them into VSP. Then VSP computes:

- the mean, $\bar{y}_{cs}$, and it's standard deviation, $s_{\bar{y}_{cs}}$, using Equations 9.1 and 9.2, respectively, in Gilbert (1987, page 107)
- the confidence interval on the mean assuming the data are normally distributed or that n' and n are large enough such that the estimated mean is normally distributed
- the estimated correlation coefficient, $\hat{\rho}$, between the two types of measurements, and the estimated standard deviation of the expensive measurements.

The correlation and standard deviation are computed so that the VSP user can evaluate if the value of those parameters that were entered into the VSP DQO dialog box are valid. If not, the new values can be entered into VSP to obtain revised values of n' and n. VSP also produces a regression plot of the inexpensive and expensive measurements so the user can graphically evaluate the linear regression assumption. Also, VSP provides a warning to the user that the computed confidence interval may be too short if n' and n are very small.

If CS is <u>not</u> cost effective, then VSP assumes simple random sampling and only the expensive measurement method will be used. VSP computes the required number of samples, n, using the iterative procedure in Gilbert (1987, page 30). Once the n expensive measurements are entered into VSP, then VSP computes the confidence interval assuming the data are normally distributed, i.e., by using the t distribution with n-1 degrees of freedom.

4.0 Test if the Mean Exceeds a Fixed Threshold Value

Suppose the sampling objective is to estimate the mean and conduct a one-sample test of the null hypothesis that the mean exceeds a fixed threshold value. The methodology for computing n' and n needed for the test has recently been developed by the authors and is currently being coded into VSP. The VSP user will access the dialog box for this methodology by clicking **Sampling Goals > Compare Average to Fixed Threshold > Can Assume Data will be Normally Distributed > Collaborative Sampling > Simple Random Sampling** or **Systematic Grid Sampling.**

First, the VSP user inputs the following data quality objective into the VSP dialog box: the null hypothesis of interest ("true mean ≥ threshold value," or "true mean ≤ threshold value," the tolerable probability, α, that the test will falsely reject the null hypothesis, the tolerable probability, β, that the test will falsely accept the null hypothesis, the width of the gray region, Δ, in the Decision Performance Goal Diagram, the expected total standard deviation of the set of expensive measurements, $\sigma_{total,ex}$, the expected correlation, ρ, between the inexpensive and expensive measurements, and the measurements costs C_{ex} and C_{inex}.

Then VSP uses Equation 1 above to determine if CS is cost effective relative to simple random sampling.

If CS is cost effective, then VSP computes n' and n using the following equations, which were derived by the authors using the method of proof used in Appendix A in EPA (2000b)

$$n' = \left[\frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} z^2_{1-\alpha} \right] \rho\left(\sqrt{R\left(1-\rho^2\right)} + \rho\right)$$

$$n = \left[\frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} z^2_{1-\alpha} \right] \left[1 - \rho^2 + \rho\sqrt{\frac{\left(1-\rho^2\right)}{R}} \right]$$

After the n' inexpensive and n expensive measurements are obtained and entered into VSP, then VSP computes

- The mean, $\bar{y}_{cs}$, and it's variance, $s^2_{\bar{y}_{cs}}$, using Equations 9.1 and 9.2, respectively, in Gilbert (1987, page 107)
- The correlation coefficient between the two types of measurements and the standard deviation of the expensive measurement, and
- A one-sample Z test of the null hypothesis

If the VSP user selected the null hypothesis to be "true mean exceeds the fixed threshold value," then the Z test is conducted by computing

$$Z = \frac{\bar{y}_{cx} - ThresholdValue}{s_{\bar{y}_{cs}}}$$

and rejecting the null hypothesis if $Z \le -z_{1-\alpha}$, where $z_{1-\alpha}$ is the (1-α)th percentile of the standard normal distribution. The Z test is used instead of the t test because the most appropriate method for determining the degrees of freedom for the t test for the CS design has not yet been determined.

VSP also constructs a regression plot of the two types of measurements, and if both n' and n are small, warns the user that the test result may not be reliable. Finally, VSP automatically reduces the value of ρ entered in the dialog box by 0.10 units (say from 0.80 specified in the dialog box down to 0.70) and re-computes n' and n. This permits the VSP user to see how n' and n change if the original value of ρ was too large by 0.10. VSP also conducts a sensitivity analysis to determine how n' and n are affected by changing the DQO input parameters. This sensitivity analysis is provided in the automatically-generated design report. This report may be inserted in a Quality Assurance Project Plan or similar project documents.

If CS is <u>not</u> cost effective, then VSP computes the number of expensive measurements, n, needed to test the null hypothesis using the following equation (derived in Appendix A of EPA 2000b), which is suitable if simple random sampling is used:

$$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} Z^2_{1-\alpha}$$

After the n expensive measurements are obtained, VSP computes the mean and its standard deviation using standard statistical formulas appropriate for simple random sampling (not CS). Finally, VSP performs a one-sample Z test and reports whether the null hypothesis can be rejected at the α-significance level.

An example of some VSP output when CS is cost efficient is shown in Figure 1.0, which shows the VSP dialog box and DQO inputs, the resulting number of samples (n' and n) computed by VSP, and the sampling locations placed on the map of the site.

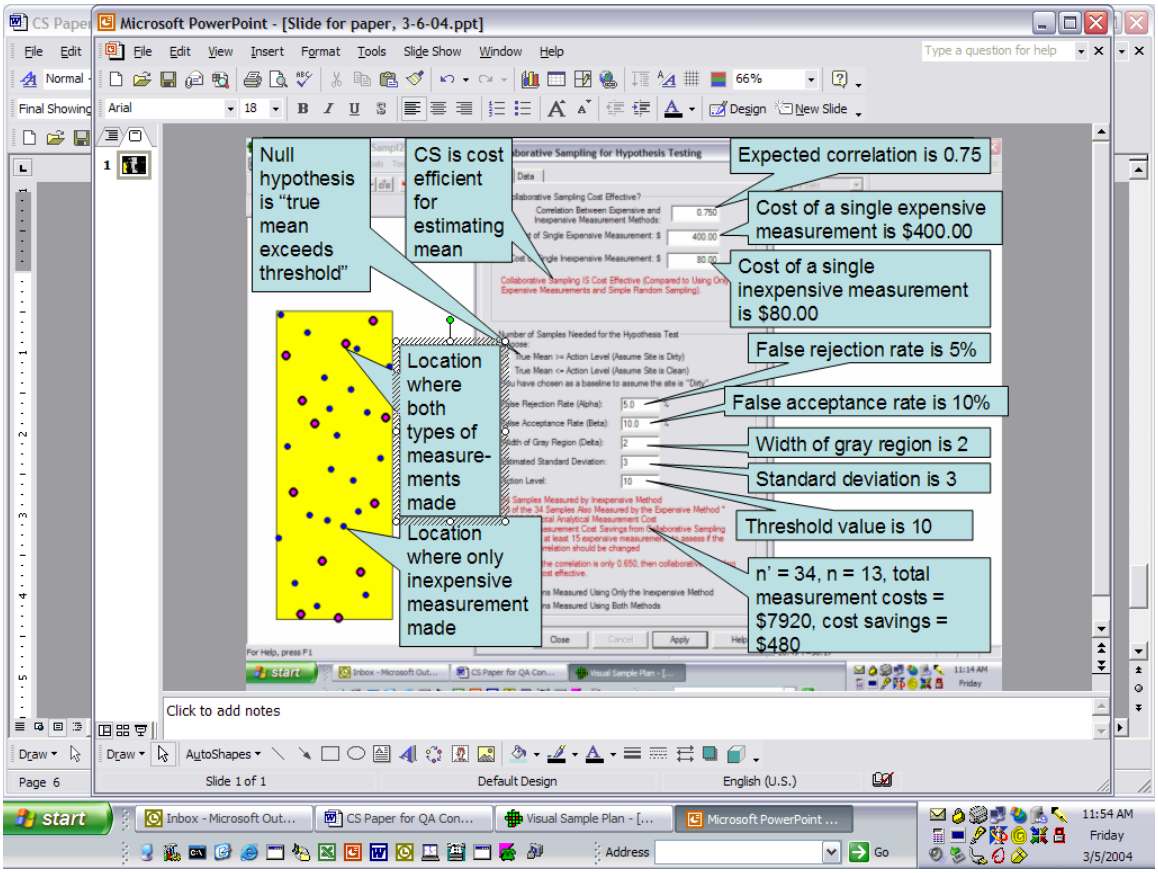


Figure 1. Example VSP Dialog Box and Map for Hypothesis Testing

5.0 References

EPA 2000a. *Guidance for Data Quality Assessment, Practical Methods for Data Analysis, EPA QA/G-9*, EPA/600/R-96/084, U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

EPA 2000b. *Guidance for the Data Quality Objectives Process*, EPA QA/G-4, EPA/600/R-96/055, U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

EPA 2002. *Guidance on Choosing a Sampling Design for Environmental Data Collection , EPA QA/G-5S,* EPA/240/R-02/005, U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*, Wiley & Sons, New York.

ETD QA Core Team: An Eloquent Solution to a Complex Problem

Thomas J. Hughes, QA and Records Manager
Experimental Toxicology Division (ETD)
National Health and Environmental Effects Research Laboratory (NHEERL)
U.S. EPA, Office of Research and Development
RTP, NC 27709

ETD is the largest health division in NHEERL, and the 120 scientists in ETD conduct more than 50 research studies each year. Many of the studies in ETD are high profile studies through interagency agreements with EPA Regions (World Trade Center Dust Study), the U.S. Air Force (Dioxin/Diabetes Study) and the EPA Office of Water (Disinfection By-Products Study in Drinking Water). Therefore, a large number of technical system reviews (TSRs), surveillances and audits (more than 30) must be conducted in ETD each year. Consequently, Tom Hughes, the ETD QA Manager, obtained permission from the Branch Chiefs and Division Director in 2003 to form an eight-member ETD QA Core Team. Team Members assisted Tom in TSRs, audits and surveillances. They were trained in QA procedures, attended national EPA (New Orleans) and ORD (Ada) QA Meetings, and assisted Tom in writing QA reports. The ETD QA Core Team for 2003 included Mette Schladweiler, Lenny Walsh and Najwa Coates from the Pulmonary Toxicology Branch (PTB), Don Doerfler from the Immunotoxicology Branch (ITB), and Karen Herbin-Davis, Tracey Ross, Carol Mitchell and Brenda Edwards from the Pharmacokinetics Branch (PKB). Participation on the ETD QA Core Team enabled Team Members to have more accurate research records, and enabled them to more fully assist their PIs when their labs were audited. In addition, they became QA ambassadors in their respective research Branches. Team Members assisted Tom for a maximum of two weeks per year, but only when their assistance did not interfere with their laboratory research. A large QA Team allowed the QA Manager the flexibility to avoid conflict-of-interests during audits, and to use alternative Team Members on audits when selected Members were not available. Membership on the ETD QA Core Team will be reviewed each year, but it is anticipated that Team Members will serve a minimum of three years. Membership on the ETD QA Core Team is viewed by Division Management as a career-enhancing opportunity. Exceptional performances by Team Members were recognized by Management in 2003 with two Team Awards. The ETD QA Core Team was recognized as a best practice by Brenda Culpepper, the DQA for NHEERL, during a QMSA in 2003. The ETD QA Core Team is an eloquent solution which has solved many QA challenges in ETD. It has greatly improved research records in the entire Division, has provided a trained QA Team to conduct required Agency QA activities on ETD research studies, and has instilled an understanding of the usefulness of QA throughout the Division. This is an abstract for presentation which has been reviewed by the U.S. EPA; views expressed do not necessarily represent EPA policy.

Quality Assurance in Research Laboratories: Rules and Reason

Ron Rogers, U.S. EPA

The progression of QA policies and their interpretations, at all EPA levels, has at times been troublesome to some scientists and QA professionals in EPA's Office of Research and Development, suggesting a need for more open discussions among all stakeholders than routinely occurs. One QA size does not fit all.

Additional emphasis has been added to this topic by the recent Agency Policy Directive for Ensuring the Competency of EPA Laboratories, which was developed by the Forum on Environmental Measurements as directed by the Science Policy Council. That proposed policy is an exquisite example of an action taken without adequate input from several primary stakeholder groups, in this case the QA community, managers, and research scientists of the National Laboratories in the Office of Research and Development (ORD). A great deal of consternation has resulted from the after-the-fact manner in which those stakeholders learned of that policy. Unfortunately, that recent episode is but one of several such QA policy decisions and interpretations that have earned uneven acceptance across the Agency and particularly within ORD over the past two decades. At the core of each such situation are two constants:
(a) a perceived lack of understanding by the QA policy decision makers of the exploratory nature of much of the research that is carried out within ORD, and
(b) a tendency toward all-or-none views by both the "enforcers" of those policies and by the scientists who are expected to comply with them.

Research vs Testing:
The scientific data that are collected by or for EPA, which are very broadly described as "environmental measurements" in each version of EPA Order 5360.1, the EPA QA mandate, typically fall into two general categories, which are most often termed "research" and "testing". In many, perhaps most, cases where QA policies have been unevenly accepted across ORD, an apparent failure to understand, acknowledge, and accommodate the important distinction between those two categories of data collection is at the very core of the problem.

> "Research" is often hypothesis driven, but not always. It is also often characterized as creative, evolving, dynamic exploration and discovery – perhaps changing course repeatedly in response to observations. The "omics" technologies that are changing the landscape of life science research are good examples. Even now, a few years and many terabytes of data into those areas of science, research studies still are usually described as "discovery science", with each subsequent experiment substantially framed by the observations of the former.
>
> On the other hand, "testing" is generally carried out under strict control, more frequently by a contractor with funding tied to the work plan, using well defined and validated procedures that are approved by a sponsor for use in the project or study. Intramural "research" funding is ordinarily provided on an annual basis to one or more investigators for studies in a research area rather than for a specific project. Portions of ORD are still

heavily dependent on contractor support for projects. A portion of ORD intramural studies employs assays and procedures have been completely reduced to routine, while others designed to provide direct support to Agency regulatory activities may have in place such rigidly prescribed objectives and operating procedures that those studies may also be viewed as "testing".

In "testing" one of the primary goals is to avoid the unexpected; that is one of the major justifications for a thoroughly detailed study protocol and the rigid requirement for adherence with that protocol and documentation of any deviations from it. In contrast, although they need not invite the unexpected in order to be sufficiently challenged by their work, many of the most important discoveries in recent times and over the ages have resulted from a well trained and experienced researcher recognizing, evaluating, and pursuing an unexpected observation or event. That creative ability to adjust and follow the course(s) that the research itself suggests is at least as important as designing and/or carefully following an established protocol in order to avoid the unexpected.

Rules Application to All-or-None:
Every version of EPA Order 5360.1 and the EPA QA Manual that supports its implementation have been abundantly clear in the Agency's intent that any data that are collected by or for the Agency must be able to withstand the scrutiny that can be anticipated should its intended use include support of regulatory decisions. That support for Agency regulatory activities is, in fact, the key distinction between QA categories 1 & 2 versus 3 & 4 as ORD applies a graded approach to QA requirements. Nevertheless, the full scope of ORD research is at least superficially included in that QA mandate, and the perspective of the Agency's Quality Staff (QS) is that the fundamental mandate applies without exception to ALL. That is rightly so, for ORD research, even the most basic and exploratory, is fundamentally applicable to EPA's efforts to identify, evaluate, and mitigate the risk associated with exposures to a host of environmental agents, and the objective defensibility of ORD research data is very important. The QS interpretation and application of portions of the Order may, however, be at issue for some ORD studies.

The EPA QA Order and Manual focus conspicuously and overwhelmingly on QA/QC issues for field studies (e.g., number and distribution of "environmental samples" to be collected and analyzed; the use of "matrix", "spiked", and other analytical control samples; etc.). That's understandable and reasonable, given that the vast majority of EPA data collecting resources are directed toward those Regional and Program Office activities of monitoring and analysis in support of site-specific decisions. Several of ORD's research programs include field studies, but the nature and objectives are often more fundamental that the regulatory objectives of Regional and Program Office decision makers. However, as a result of that language bias, the QA mandate's relevance is not all that clear to researchers in several other areas. For example, field study terminology is not familiar to researchers who are attempting to devise and refine new in vitro or in vivo assays that may permit one to better understand the changes at a molecular, cellular, organ, or organism level that underlie reported effects of environmental exposures on vision, memory, immunocompetence, tumor development, birth defects, etc. There are no such things as spiked vision controls, nor a sample matrix for memory measurements, but there are numerous controls (e.g., positive control, negative control, vehicle control, cage control, sham exposure control), that are used to help dissect observations in order to identify those

components attributable to the environmental agent(s) under investigation. ***The terminology may be different but the principles are the same!*** Also important to this discussion, there are no NIST-traceable standard reference rodents or fish that can be used to represent "true" values for toxicology measurements although transgenic animals represent a step in that direction

The extreme focus that has been placed, especially in recent years, on the development and institutional approval of a QA Project Plan (QAPP), or equivalent, for each study before any data collection is initiated has become the dominant policy feature, sending a very loud message that investigators must obtain official written, study-specific permission before proceeding with their research. That is interpreted in the ORD context to mean that ALL studies are to be conducted as if they can be described as "testing" or are being performed under contract. That issue may have been taken to the extreme recently when an NHEERL researcher was reportedly advised by QS that even a literature search should not be initiated until an approved QAPP is in place.

When faced with a mandate that does not appear to apply to their particular activities, many will choose to dismiss the mandate as irrelevant and thereby ignore any valid issues and principles that may underlie it (e.g., records, data verification, error, uncertainty). QA is not n/a for research, far from it – some would argue that QA is even more important in research for the very same reasons that others suggest it is n/a, that is, the lack of knowledge of "true"values, lack of standards, lack of validated methods, etc. It is perhaps worth noting that ORD Laboratories had developed QA programs several years before the initial Agency QA mandate in 1979. However, when the relevance of a particular Agency requirement is not made clear, the ORD research audience is at best – confused, and at worst – reluctant and perhaps resistant. That has been a predictable result of attempts to impose "testing" rules on "research" studies, leaving ORD QA staff with an unnecessarily difficult challenge.

Fortunately, the QA Order and Manual contain other requirements (e.g., periodic updates of QAPPs) and allowances (e.g., the flexible graded approach) that are logical and practicable for research, although grossly under emphasized in comparison to the QAPP-before-data rule. Having made the distinction between "research" and "testing" sufficiently clear (hopefully), attention can be turned to specific issues encountered by ORD QA Managers. Building on the opportunities that the graded approach, periodic updates, and other flexible interpretations permit, we should be able to achieve improvements in our interactions with QS and with our researchers. From the ORD perspective, because that is where we directly see the need to ensure that appropriate QA policies and practices are in place for application to our research efforts, the ORD QA professionals, with QS concurrence, need to define (and heed) appropriate policy application and exceptions in order to permit reason and experience to control QA interpretations and implementation in ORD research programs. ORD QA staff have recognized and should be permitted to focus QA efforts on the QA issues that are seen to present greater relevance and potential benefit to the quality and defensibility of our research data. Otherwise, we risk the erosion of the credibility and trust that ORD QA Managers have gained in the past decade from the scientists that we support. That trust has been gained in part by demonstrating our confidence that our scientists have no incentive to fabricate data or ignore errors – their professional reputations are based on those data and routinely on the ability of others to replicate and validate their results. The lack of trust that resulted from the contractor misdeeds in the '70s

that led to the EPA QA mandate may or may not still be appropriate for data collection contractors, but it is certainly not a broad concern for in-house EPA researchers.

The so-called "graded approach" that EPA permits in QA planning was, it seems, primarily designed to accommodate projects or studies that differ in scale/scope, goals (i.e., anticipated data use), progression phase, etc. It should be expanded to more clearly accommodate the divergent quality issues that exist between "research" and "testing" studies. Achieving that flexibility while remaining faithful to the Agency's QA mandate is not a trivial effort, but in the interest of the quality of ORD science, we must continue to pursue it! There are valid reasons (again, most originating from the performance of those few corrupt contractors in the '70s) for the rules laid out in the EPA QA Manual, but not all of those reasons apply to all data collection activities. Breaking the major QA Manual topics down we can look at some potential benefits and costs of compliance for ORD.

Quality System/Quality Management Plans: Documenting organizational QA policies and procedures and the roles and responsibilities of individuals at every level in the implementation of those policies and procedures is essential if the Quality System is to succeed in supporting the collection of data and defense of those data. The Labs, Centers, and even Divisions within the Labs have done so, but the ORD QA community has been unable to reach consensus as to whether there is a need or benefit to defining a Quality System at the ORD level.

QA Project Plans (QAPP): The QAPP (or equivalent) requirement is a prime example of where flexibility and improvement is vital, well within reason, and within reach. Although explicitly required by QS before data collection is begun, preparing such a planning document at a meaningful level of QA/QC detail before preliminary research is underway and preliminary data have been gathered is often difficult and rather futile, especially for many QA category 3 or 4 studies. Attempts within ORD to achieve compliance with that QS requirement have taken various courses, with varying results. Until approximately a decade ago, several ORD Laboratories were heavily dependent on contractors for our QA programs, and the result was in some cases compliance on paper but a less impressive reality. At about that same time, QS (then known as the Quality Assurance Management Staff) began creating the Agency collection of requirement and guidelines documents that now exists, and the EPA Office of Health Research and Health Effects Research Laboratory (HERL) did similarly, developing QA guidance tailored to the HERL audience. In recognition of the broad range of study types across HERL, that guidance was issued as just that – guidance – rather than policy. Dr. Karen Gold (RTI, my QA mentor) and I set about the task of bringing into focus the concept of a QAPP or equivalent (which we termed "Intramural Research Protocols") at the research study level. Doing so required that we define "study" in a more specific way – a body of work for which one can predict a beginning and end; the level at which specific QC activities can be applied (and thus detailed in a QA plan), and the level for which a complete collection of records should be assembled that can be reviewed and understood. A major focus has since that time been on the study record as the primary means for demonstrating and defending the quality of the results, but always encouraging sufficient attention to statistical design, data verification, and data management during the planning process. The EPA QA Managers who replaced that QA support contract in 1995 have made great progress in promoting compliance with the QAPP requirement and improvements in the creation and management of study records.

Unlike the technical reports that other parts of the Agency may prepare, NHEERL's primary work products are journal articles, and the type of QA information that is needed to defend reported results is not permitted by those journals. If the QAPP requirement were to be interpreted and applied primarily as a records management tool rather than as a permission slip (which is usually not required in the in-house research environment), the resulting document, or as is often the case, series of documents capturing the evolving progression of an ORD research study would be an extremely useful element of the completed study record. Given that the study record should include everything that is needed to recreate the study or reanalyze the data, including raw and all other forms of data plus any manuscripts reporting the results, it is the ultimate product of the study. The defensibility of a research study will rely on the study record, much more so than on the existence of an approved QAPP that predates any collected data A QAPP, updated as frequently as is warranted (perhaps in an electronic system that simplifies and streamlines that process), provides the framework and context for that study record.

In a somewhat similar fashion, certain ORD organizations have achieved compliance by a step-wise approach, permitting approval of a *de minimis* QAPP for a set period of time (e.g., one year), during which time the study may be initiated, but requiring a full QAPP, consistent with EPA QA/R-5, at the end of that period. Other organizations have made the QAPP primarily a budget and resource management tool, with excellent results with regards to compliance but uneven and/or as yet uncertain results with regards to QA objectives and records management.

ORD has in recent years moved toward increased use of research teams to address issues. In such cases, the QAPP, whether written as a linked collection of sub-project QAPPs or as a single overarching QAPP, can be an extremely valuable tool for achieving consensus on the course of research, ensuring a common understanding of the project objectives and quality criteria, and for communicating to all parties any requirements for documentation and labeling that may be necessary to ensure that a study record can be assembled that can endure and be easily understood. In these large team projects, it is often the case that the planning and implementation evolve over several months. For example, what certain investigators choose to do may depend on whether they are able to develop suitable methods or on the results that others investigator achieve with their experiments, thus staggering important planning decisions. In such cases, it is quite difficult to fully complete an R-5 compliant QAPP before preliminary data are collected, but also especially important that an R-5 level of detail be incorporated as soon as that information becomes available and those planning decisions are made.

Quality Assessments: QA reviews should be regularly conducted at both the Quality System level and at the project or study level. The use of the term "audit" is probably appropriate for contract work or "testing" where there are policies and procedures to which strict adherence is demanded, but not for assessing the implementation of a Quality System or research project where flexibility is essential to fulfilling the mission and goals is quite inappropriate. "Audit" universally suggests a determination of whether an act or item is correct, and thus requires a standard against which that determination can be made. "Auditing" a research program or individual study risks discrediting the "auditors", for they may be tempted to inappropriately establish standards where none exist, or apply misguided interpretations of rules or guidelines. For example, NHEERL's set of QA **guidelines** documents, mentioned earlier, were explicitly

and deliberately created as **guidelines** rather than requirements, because the broad and varied range of disciplines and studies within HERL demanded that flexibility. However, they were misinterpreted in QS as requirements, and applied as the basis for "audit" findings.

Various types of reviews are routinely conducted in ORD Labs, and are useful, occasionally resulting in detecting errors and finding opportunities for improving processes on which the quality and defensibility of researchers' data and other study records depend. However, the benefit that can be achieved by developing and refining tools that simplify the documentation of the entire research process – project planning and implementation; data collection, verification, and management – will certainly exceed any benefit from the external QA reviews, not to mention that due to an approved pre-study QAPP. As we face major challenges in migrating toward new research technologies and increased electronic data and records management, we have both a need and an opportunity to build appropriate QA rules and reason into those processes. NHEERL and others in ORD have begun the use of NuGenesis, a Scientific Data Management System, and numerous Lotus Notes system tools and programmed applications in an attempt to better and more easily develop, track, and store all aspects of a study record. The openness of such an e-record approach will greatly enhance opportunities for peer and QA input, demanding the utmost integrity in the planning and implementation of ORD research, and the technology will also enable the easy demonstration of that integrity. Input by the ORD researchers, ORD QA professionals, and EPA Quality Staff in the design and completion of those undertakings will be essential.

ORD scientists and QA professionals have in the past decade made tremendous progress toward fulfilling the principles and objectives of the EPA Order 5360.1. That has come primarily from the willingness of ORD's Laboratory and Division managers, primarily the latter, to embrace and support the "good science" principles that underlie the EPA QA mandate, plus the ORD QA professionals' ability to adapt Agency requirements and guidelines to the reality of ORD research. To continue that progress, we will need even more support from our managers, and that is likely to depend on how logical, beneficial, and defensible our QA work products (i.e., requirements and guidance) are. We can help our cause if we exercise the trust that is warranted in our ORD programs and make the so called "graded approach" truly graded by leaving to the discretion of ORD division Quality Management Systems the decisions and policies about how the EPA QA Order will be implemented in ORD – for example, the types of studies for which a QA plan will be required before preliminary data may be collected, and how the QAPP will be used by the organization. We must also be willing to step forward to ensure that the ORD stakeholder group is sufficiently represented whenever QA (or other) policies and procedures are at issue. Imposing ill-conceived requirements only does harm, and ORD's dedicated Sisyphusian QA professionals need no more mass added to our boulders.

The views expressed in this paper are those of the author, and do not necessarily reflect US EPA policy nor the views of the author's ORD QA colleagues.